Compressive Link Acquisition in Multiuser Communications

Xiao Li†, Andrea Rueetschi†, Anna Scaglione† and Yonina C. Eldar*

***Abstract*—An important receiver operation is to detect the presence of specific preamble signals with unknown delays and in the presence of scattering, Doppler effects and carrier offsets. This task, referred to as "link acquisition", is typically a sequential search over the transmitted signal space. Recently, many authors have suggested applying sparse recovery algorithms in the context of similar estimation or detection problems. These works typically focus on the benefits of sparse recovery, but not generally on the cost reduction brought by compressive sensing. Thus, our goal is to examine the trade-off in complexity and performance that is possible when using sparse recovery. To do so, we consider a sequential Compressive Sparsity-Aware (C-SA) acquisition scheme, where a Compressive Multi-channel Sampling (CMS) module is followed by a Sparsity Regularized (SR) Likelihood Ratio Test (LRT) receiver.**

Our C-SA acquisition scheme borrows insights from the models studied in the context of sub-Nyquist sampling, where a minimal amount of samples is used to reconstruct signals with Finite Rate of Innovation (FRI). In particular, we propose an A/D conversion front-end that maximizes the Kullback-Leibler distances among the hypotheses of the SR-LRT performed on the samples. We compare the proposed acquisition scheme vis-à-vis conventional alternatives with relatively low computational cost, such as the Matched Filter (MF), in terms of performance and complexity. Our experiments suggest that one can use the proposed C-SA scheme to scale down the receiver implementation cost, with greater flexibility than conventional MF architectures. However, we find that they both have overall complexities that scale linearly with the search space and that compressive measurements used in the SR-LRT at low SNR lead to a performance loss, as one could expect given that they use less observations. At high SNR, on the other hand, the SR-LRT has better performance in spite of the compression.

***Index Terms*—Multiuser communications, compressed sensing, sparse recovery, detection and estimation.**

I. Introduction

One of the critical receiver tasks in a multiuser scenario, referred to as *link acquisition*, is that of detecting the presence of signals, and identifying the *link parameters* (e.g., delays, Doppler) of an *unknown subset* $\mathcal{I}$ of active sources out of I possible ones. Similar to [1], [2], we consider the case in which the set of active users $\mathcal{I}$ transmit known and distinct training preambles $\phi_i(t)$, $i \in \mathcal{I}$. Usually these preambles are fairly long, so that their energy at the receiver can rise above the noise. In the initial phase the receiver is completely agnostic about sources surrounding it and it tests the sensed signal $x(t)$ until it detects the presence of such signals, to establish the active links. This needs to be done accumulating observations and repeating the test sequentially. The acquired link information is essential for identifying the basic features of the active signal spaces, so that the receiver can determine if it can decode the data after the training phase [3], [4] and refine the link parameter estimates using mid-ambles and decoded data. The term "link acquisition" is equivalent to resolving the received signal space, which is characterized by the propagation delays and the Doppler frequencies.

A. Related Works on Link Acquisition of Multiuser Signals

We can classify the algorithms that are used for link acquisition into two main groups. The first category acquires a *sufficient statistic* by directly sampling $x(t)$ at (or above) the Nyquist rate. The likelihood function is then exploited to detect the presence of signals and determine the link parameters in the model given the set of active users $\mathcal{I}$. We refer to such techniques as Direct Sampling (DS) methods (e.g. [5]–[8]).

A second approach, referred to as the Matched Filtering (MF) [4], [9], [10], facilitates the search of both the *active set* $\mathcal{I}$ and link parameters by comparing the filtered outputs of the signal $x(t)$ from a bank of filters $\phi_i(t-\tau)e^{-\mathrm{i}\omega t}$, each matching a sufficiently wide collection of points in the full parameter set $\mathcal{T} \times \mathcal{F}$ where $\tau \in \mathcal{T}$ and $\omega \in \mathcal{F}$ are the delay and Doppler spread respectively. MF is a prevalent choice in hardware implementations because of its simplicity. The MF approach can be implemented in two ways: in the digital domain as a DS method, where samples are projected onto the sampled version of $\phi_i(t-\tau)e^{-\mathrm{i}\omega t}$, or directly in hardware, as a multi-channel sampling structure, perfoming the projections onto the templates $\phi_i(t-\tau)e^{-\mathrm{i}\omega t}$ via analog filters. Specific details on these algorithm architectures are provided in Section III.

Classical algorithms take little advantage of the low dimensionality of the received signal space in storing the observations that are processed or to improve the detection performance. Recently, there have been advances in exploiting *sparsity*, or the low dimensionality of the signal space, to improve receiver performance. One class of papers suggests using sparse signal recovery for the purpose of either detection or estimation. For instance, assuming that the signal is present, [1], [2], [4], [11] deal with identification of the active users and/or estimation of signal parameters by creating a dictionary from the known templates $\phi_i(t)$ and viewing the signal $x(t)$ as a sparse linear combination of these element templates inside the dictionary. Without knowledge of signal presence within a specific observation window, the proposed detection

† The authors are with the Department of Electrical and Computer Engineering, University of California, Davis 95616, USA, email : {eceli,andrear,ascaglione}@ucdavis.edu

* Y. C. Eldar is with Department of Electrical Engineering, Technion, Israel Institute of Technology, Haifa 32000, Israel, email : yonina@ee.technion.ac.il

schemes in [12]–[16] use generic compressed measurements to detect the presence of certain signals, starting from an abstract discrete model. We call this class of methods Sparsity-Aware (SA). In these papers, Doppler shifts and delays are not explicitly considered and the discrete observations are treated independently as a single snapshot, upon which SA algorithms are applied.

B. Multiuser Signals with Finite Rate of Innovation (FRI)

What is often neglected in existing SA approaches is the acquisition of informative low rate discrete samples from the analog domain. As we mentioned, preamble sequences are usually fairly long and the receiver needs to sample the signal $x(t)$ at a fast rate and store them prior to processing. This can become a bottleneck in designing preambles so that they have the appropriate energy.

Reducing the sampling rate and the associated storage incurred at the A/D front-end is mostly the concern of another broad class of papers [17]–[21] on signals with a Finite Rate of Innovation (FRI) [22]. In general, an FRI model has a *sparse* parametric representation. Given the preamble $\phi_i(t)$ for each active user $i \in \mathcal{I}$ traveling through R multipath channels, the class of signals $x(t)$ lies in a subspace with no more than $|\mathcal{I}|R$ dimensions, where each dimension has three unknowns (e.g., delay, Doppler, channel coefficient), irrespective of its bandwidth and duration.

The premier objective of FRI sampling architectures is A/D conversion at sub-Nyquist rates (i.e. deterministic signal reconstruction). This objective is fundamentally different from what is of practical interest in link acquisition, which is to perform statistical inference. In this paper, we wish to harness similar complexity gains as in the FRI literature, while mitigating the detection performance losses that arise in the presence of noise due to the reduced number of observations. To this aim, we formulate the link acquisition problem as a Sparsity Regularized (SR) Likelihood Ratio Test (LRT) that uses compressive samplers maximizing the average Kullback-Leibler (KL) distance among the hypotheses in the test. We refer to the compressive samplers designed in this paper as the Compressive Multichannel Sampling (CMS) architecture and the proposed link acquisition scheme as the Compressive Sparsity-Aware (C-SA) scheme. More specifically, we discuss in this paper

1) a unified low-rate CMS architecture for the C-SA acquisition scheme using the proposed SR-LRT;
2) the sequential SR-LRT that jointly detects signal presence and recovers the active users with their parameters;
3) the compressive samplers of the C-SA scheme that provide the maximum average Kullback-Leibler (KL) distance of the SR-LRT;
4) the comparison of the proposed architecture with the MF approach in terms of performance, storage cost and computational complexity.

This bridges the results pertaining to sparsity-aware estimation/detection [1], [2], [4], [12]–[16], the literature on analog compressed sensing and sub-Nyquist sampling [4], [8], [10], [17], [18], [23] and FRI sampling [19], [20], [22] such that sampling and estimation/detection operations are considered jointly.

To measure the benefits of the proposed C-SA scheme over other schemes, we analyze the practical trade-off between the implementation costs spent in acquiring samples and those invested in sparse recovery. This is important to clarify the potential benefits of sub-Nyquist architectures in communication receivers in the link acquisition phase. These schemes often benefit from the denoising capabilities of SA algorithms (as well documented in [1], [2], [4], [12]–[16]) but must loose sensitivity due to the fact that they do not use sufficient statistics for the receiver inference.

The question we consider is, therefore, *what is there to gain: hardware, storage, complexity or performance?* Our numerical experiments indicate that the main advantage of the proposed scheme is that it enables the designer to find an adequate operating point for link acquisition such that processing requirements and complexity of the receiver can be reduced to an acceptable level without significantly sacrificing acquisition performance compared with the MF architecture. We also confirm numerically that the compressive samplers we propose in the CMS architecture harvest highly informative samples for the SR-LRT in terms of estimation and detection performance.

C. Notation and Paper Organization

We denote vectors and matrices by boldface lower-case and boldface upper-case symbols and the set of real (complex) numbers by $\mathbb{R}$ ($\mathbb{C}$). We denote sets by calligraphic symbols, where the intersection and the union of two sets $\mathcal{A}$ and $\mathcal{B}$ are written as $\mathcal{A} \bigcap \mathcal{B}$ and $\mathcal{A} \bigcup \mathcal{B}$ respectively. The operator $|\mathcal{A}|$ on a discrete (continuous) set takes the cardinality (measure) of the set. The magnitude of a complex number x is denoted by $|x| = \sqrt{xx^*}$, where x^* is the conjugate of the complex number x. The transpose, conjugate transpose, and inverse of a matrix $\mathbf{X}$ are denoted by $\mathbf{X}^T$, $\mathbf{X}^H$ and $\mathbf{X}^{-1}$, respectively. The inner products between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^{N\times 1}$ and between two continuous functions $f(t), g(t)$ in $L_2(\mathbb{C})$ are defined accordingly as $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=1}^{N} y_n^* x_n$ and $\langle f(t), g(t) \rangle = \int_{-\infty}^{\infty} g^*(t) f(t) \mathrm{d}t$. The $\mathbf{W}$-weighted ℓ_2-norm of a vector $\mathbf{x}$ is denoted by $\|\mathbf{x}\|_{\mathbf{W}} = \sqrt{\mathbf{x}^H \mathbf{W} \mathbf{x}}$, and the conventional ℓ_2-norm is written as $\|\mathbf{x}\|$. The L_2-norm of a continuous-time signal $f(t) \in L_2(\mathbb{C})$ is computed as $\|f(t)\| = \sqrt{\langle f(t), f(t) \rangle}$.

The paper is organized as follows. Section II introduces our observation model. We discuss related works on link acquisition in Section III. The CMS architecture for compressive acquisition is considered in Section IV. Using the compressive samples obtained from the CMS architecture, we develop the SR-LRT algorithm for C-SA link acquisition scheme in Section V. We then optimize the compressive samplers in the CMS in Section VI. Simulations demonstrating the performance are presented in Section VII. The overall cost of the proposed C-SA scheme using the CMS module is compared against conventional MF schemes in terms of computational complexity and storage in Section VIII.

II. Sequential Link Acquisition

In every communication standard, a key control sequence in the training phase is the initial preamble. The receiver models the corresponding observation by assuming that each $i \in \mathcal{I}$ from the unknown active set transmits a specific preamble $\phi_i(t)$. This transmission is followed by the mid-ambles and data frames. A common choice for such a preamble in multiuser communications is a linearly pulse modulated sequence with a chip rate $1/T$ close to the signal bandwidth and equal to the minimum Nyquist-rate

$$\phi_i(t) = \sum_{m=0}^{M-1} a_i[m] g\left(t - mT\right). \tag{1}$$

Here $g(t)$ is the pulse shaping filter (chip) and $\{a_i[m]\}_{m=1}^{M}$ is typically a long preamble sequence $M \gg 1$ for each user.

Then the observation at the receiver can be written as

$$x(t) = \sum_{i\in\mathcal{I}} \sum_{r=1}^{R} h_{i,r}\phi_i(t - t_{i,r}) e^{\mathrm{i}\omega_{i,r}t} + v(t), \tag{2}$$

where $t_{i,r}$ is the unknown propagation delay of the ith user in the rth multipath, $|\omega_{i,r}| \le \omega_{\max}$ is the Doppler frequency upper bounded by the maximum Doppler spread $\omega_{\max}$, and $h_{i,r}$ is the unknown channel fade. Without loss of generality, we assume that the maximum multipath order R is known and the noise component $v(t)$ is a white Gaussian process with $\mathbb{E}\{v(t)v^*(s)\} = \sigma^2\delta(t-s)$. Our problem is to detect the presence of the active user set $\mathcal{I}$ and the corresponding link parameters $\{h_{i,r}, t_{i,r}, \omega_{i,r}\}$ for $i \in \mathcal{I}$ and $r = 1, \cdots, R$.

Since the propagation delays $t_{i,r}$ are unknown and possibly large (infinite if there is no signal transmitted at all), the typical A/D front-end for link acquisition is sequential. The acquisition scheme produces test statistics every D units of time, where D is the shift in the time reference for detections. At every shift $t = nD$, the receiver decides whether the signal $x(t)$ is present at or after $t = nD$. For convenience, we denote $t_0 = \min\limits_{i,r} t_{i,r}$ as the delay of the first arrival path among all users. Let

$$\ell = \lfloor t_0/D \rfloor \tag{3}$$

be the value of $n = \ell$ that most likely produces a positive detection and

$$\tau_{i,r} = t_{i,r} - \ell D \tag{4}$$

be the *composite delay*. Clearly, we have $0 \le \tau_{i,r} \le \tau_{\max}$, where $\tau_{\max}$ is the composite delay spread. Note that the delay $\tau_{i,r}$ can fall anywhere within $[0, D)$ and thus $\tau_{\max} \ge D$. This allows us to express (2) equivalently and uniquely as

$$x(t) = \sum_{i\in\mathcal{I}} \sum_{r=1}^{R} h_{i,r}\phi_i(t - \ell D - \tau_{i,r}) e^{\mathrm{i}\omega_{i,r}t} + v(t). \tag{5}$$

After these considerations, it is clear that the search spaces of delays and Doppler frequencies for each shift n are respectively $\mathcal{T} \triangleq [0, \tau_{\max}]$ and $\mathcal{F} \triangleq [-\omega_{\max}, \omega_{\max}]$.

A. Goal of Link Acquisition

Note that there could be multiple values of $n \neq \ell$ that lead to valid positive detections, where for a given ℓ, the relative composite delay with respect to the nth shift would be

$$\tau_{i,r}^{(n)} = \tau_{i,r} - (n - \ell)D. \tag{6}$$

In order to single out the best reference shift, the receiver will have to compare a sequence of N_0 test statistics after the first positive detection at $n = N_\eta$, and choose the particular shift $\ell_\star$ that maximizes the likelihood ratio between the signal hypothesis and the noise hypothesis. We call $\ell_\star$ the *maximum likelihood ratio* (MLR) shift. The look-ahead horizon N_0 can be chosen considering the type of sampling kernels, the preambles $\phi_i(t)$'s, and the delay spread $\tau_{\max}$, making reasonable approximations about the durations of the signals.

Definition 1. Link acquisition *refers to*

1) *locating the MLR shift $\ell_\star$;*
2) *identifying the set of active users $\mathcal{I}$ in the $\ell_\star$th shift;*
3) *resolving the delay-Doppler pairs $\{\widehat{\tau}_{i,r}, \widehat{\omega}_{i,r}\}$ for $i \in \mathcal{I}$ and $r = 1, \cdots, R$.*

Usually, the preamble signals $\phi_i(t)$'s have large energy, so that they can rise above the receiver noise. Given that the average power is constant, the $\phi_i(t)$ typically last much longer (i.e., M is large) than subsequent mid-ambles or spreading codes that modulate data. For a typical wireless application such as GPS or IS-95/IMT-2000, transmitters continuously send out preamble sequences with length on the order of $M = 20 \times 1023$ [24] or $M = 32768$ [25], respectively. This means that in order to detect the presence of such preambles and acquire the synchronization parameters, architectures using DS or MF approaches would have to store a large amount of data to process in a sequential manner. This phase is crucial to properly initialize any channel tracking that ensues. In Section III, we provide details on the A/D architectures and the corresponding post-processing for conventional link acquisition schemes. We then present the proposed CMS architecture and the C-SA link acquisition scheme in Section IV.

III. Existing Architectures for Link Acquisition

For future use, we let the Nyquist rate of the signal $x(t)$ be $f_{\text{NYQ}} = 2\mathcal{W} + \omega_{\max}/\pi$ with $\mathcal{W}$ being the maximum single-sided bandwidth of $\phi_i(t)$, $i = 1, \cdots, I$.

A. Direct Sampling (DS)

In DS schemes, the received analog signal $x(t)$ is sampled by projecting it onto an ideal series of Dirac's deltas, every $T_s = 1/f_s \le 1/f_{\text{NYQ}}$, i.e.

$$c_{\text{DS}}[w] = \langle x(t), \delta(t - wT_s)\rangle = x\left(wT_s\right). \tag{7}$$

At the nth shift, DS schemes use the most recent W Nyquist samples for every shift $D = NT_s$

$$\mathbf{c}_{\text{DS}}[n] = \left[c_{\text{DS}}[nN], \cdots, c_{\text{DS}}[nN + (W-1)]\right]^T \tag{8}$$

to perform the detection. Based on (2), the samples $\mathbf{c}_{\text{DS}}[n]$ can be expressed as

$$\mathbf{c}_{\text{DS}}[n] = \mathbf{\Phi}_{\mathcal{J}}\left(\boldsymbol{\tau}_{\mathcal{J}}, \boldsymbol{\omega}_{\mathcal{J}}\right)\mathbf{h}_{\mathcal{J}} + \mathbf{v}[n], \tag{9}$$

where $\mathcal{J} \subseteq \{1, \cdots, I\}$ is the set of active users at the nth shift, and

$$\boldsymbol{\tau}_{\mathcal{J}} \triangleq [\cdots, \tau_{j,1}^{(n)}, \cdots, \tau_{j,R}^{(n)}, \cdots]^T \tag{10}$$

$$\boldsymbol{\omega}_{\mathcal{J}} \triangleq [\cdots, \omega_{j,1}, \cdots, \omega_{j,R}, \cdots]^T \tag{11}$$

$$\mathbf{h}_{\mathcal{J}} \triangleq [\cdots, h_{j,1}, \cdots, h_{j,R}, \cdots]^T \tag{12}$$

represent the $|\mathcal{J}|R$ residual delays, Doppler and channel coefficients corresponding to the set of users $\mathcal{J}$ in that shift. The vector $\mathbf{v}[n]$ contains the noise samples $[\mathbf{v}[n]]_w = v(nD + wT_s)$, and $\mathbf{\Phi}_{\mathcal{J}}\left(\boldsymbol{\tau}_{\mathcal{J}}, \boldsymbol{\omega}_{\mathcal{J}}\right)$ is a $W \times |\mathcal{J}|R$ sub-matrix of the complete $W \times IR$ matrix $\mathbf{\Phi}\left(\boldsymbol{\tau}, \boldsymbol{\omega}\right)$, from which we extract columns $j \in \mathcal{J}$. The full matrix $\mathbf{\Phi}\left(\boldsymbol{\tau}, \boldsymbol{\omega}\right)$ is defined by

$$\left[\mathbf{\Phi}\left(\boldsymbol{\tau}, \boldsymbol{\omega}\right)\right]_{w,(i-1)R+r} \triangleq \phi_i\left(wT_s - L_gT - \tau_{i,r}\right) e^{\mathrm{i}\omega_{i,r}wT_s},$$

where L_g is the number of truncated side-lobes of $g(t)$ on one side.

Using the nth shift of observations at $t = nD$, link acquisition amounts to performing the following multiple composite hypothesis test

$$\begin{aligned}
&\mathcal{H}_{\mathcal{J}} : \mathbf{c}_{\text{DS}}[n] = \mathbf{\Phi}_{\mathcal{J}}\left(\boldsymbol{\tau}_{\mathcal{J}}, \boldsymbol{\omega}_{\mathcal{J}}\right)\mathbf{h}_{\mathcal{J}} + \mathbf{v}[n],\\
&\mathcal{H}_{\varnothing} : \mathbf{c}_{\text{DS}}[n] = \mathbf{v}[n],
\end{aligned}$$

with unknown parameters $\mathcal{J}$, $\boldsymbol{\tau}_{\mathcal{J}}$, $\boldsymbol{\omega}_{\mathcal{J}}$ and $\mathbf{h}_{\mathcal{J}}$.

The Generalized Likelihood Ratio Test (GLRT) is then typically used. This test requires solving the non-linear least squares estimation (NLLSE)

$$\left\{\widehat{\mathcal{I}}, \widehat{\boldsymbol{\tau}}_{\widehat{\mathcal{I}}}, \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{I}}}, \widehat{\mathbf{h}}_{\widehat{\mathcal{I}}}\right\} = \underset{\mathcal{J}, \boldsymbol{\tau}_{\mathcal{J}}, \boldsymbol{\omega}_{\mathcal{J}}, \mathbf{h}_{\mathcal{J}}}{\arg\min} \left\|\mathbf{c}_{\text{DS}}[n] - \mathbf{\Phi}_{\mathcal{J}}\left(\boldsymbol{\tau}_{\mathcal{J}}, \boldsymbol{\omega}_{\mathcal{J}}\right)\mathbf{h}_{\mathcal{J}}\right\|,$$

over all possible $\mathcal{J}$, $(\boldsymbol{\tau}_{\mathcal{J}}, \boldsymbol{\omega}_{\mathcal{J}}) \in \mathcal{T}^{|\mathcal{J}|} \times \mathcal{F}^{|\mathcal{J}|}$, $\mathbf{h}_{\mathcal{J}} \in \mathbb{C}^{|\mathcal{J}|}$ to compute the generalized likelihood ratio

$$\eta_{\text{DS}}(n) = \frac{\mathbb{P}\left(\mathcal{H}_{\widehat{\mathcal{I}}}\right)}{\mathbb{P}\left(\mathcal{H}_{\varnothing}\right)} \tag{13}$$

with estimates $\left\{\widehat{\mathcal{I}}, \widehat{\boldsymbol{\tau}}_{\widehat{\mathcal{I}}}, \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{I}}}, \widehat{\mathbf{h}}_{\widehat{\mathcal{I}}}\right\}$ obtained at every shift $t = nD$. The expression of the generalized likelihood ratio is given in [26] for cases when the noise variance σ^2 of $\mathbf{v}[n]$ is known and unknown. Using the corresponding ratio as test statistics, the receiver checks if the test statistic satisfies $\eta_{\text{DS}}(n) \geq \eta_0$ for some properly chosen threshold $\eta_0 > 1$. Without loss of generality, we consider the most general case where σ^2 is unknown. In this case, the generalized likelihood ratio has the following expression

$$\eta_{\text{DS}}(n) = \frac{\mathbb{P}\left(\mathcal{H}_{\widehat{\mathcal{I}}}\right)}{\mathbb{P}\left(\mathcal{H}_{\varnothing}\right)} = \frac{\left\|\mathbf{c}_{\text{DS}}[n]\right\|^{2W}}{\left\|\mathbf{c}_{\text{DS}}[n] - \mathbf{\Phi}_{\widehat{\mathcal{I}}}\left(\widehat{\boldsymbol{\tau}}_{\widehat{\mathcal{I}}}, \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{I}}}\right)\widehat{\mathbf{h}}_{\widehat{\mathcal{I}}}\right\|^{2W}}. \tag{14}$$

Denote the first shift that passes the GLRT as $N_\eta \triangleq \min\left\{\underset{n}{\arg}\ \eta_{\text{DS}}(n) \geq \eta_0\right\}$, then the MLR shift $\ell_\star$ is given by

$$\ell_\star = \underset{n}{\arg\max}\ \eta_{\text{DS}}(n), \quad n = N_\eta, \cdots, N_\eta + N_0. \tag{15}$$

Obviously, the test described above is intractable in general, since there are 2^I hypotheses at each shift $t = nD$ to explore, and for each of them, there is an NLLSE problem to solve. Therefore in practice, DS acquisition schemes either deal with the known user case $\mathcal{J} = \mathcal{I}$ or assume the full set $\mathcal{J} = \{1, \cdots, I\}$ during detection, followed by NLLSE for that specific user set.

When the set of active users $\mathcal{I}$ is unknown, alternatives are *Matched Filtering* (MF) and *Sparsity-Aware* (SA) approaches. The C-SA scheme in this paper is an instance of the SA technique, which performs sequential detection and estimation using sub-Nyquist samples from the proposed CMS architecture. We next describe the MF approach and then the SA method.

B. Matched Filtering (MF)

The MF receiver is a widely used architecture in practice because of its ease in finding the active parameters by simply observing the filtered outputs of the MF filterbank, which is constructed from the *MF templates* $\phi_i(t)e^{\mathrm{i}k\Delta\omega t}$ for some i and k. Clearly, the size of the filterbank has to be finite. Therefore, it is usually assumed that $\tau_{i,r} \approx q_{i,r}\Delta\tau$ and $\omega_{i,r} \approx k_{i,r}\Delta\omega$ for some integers $q_{i,r}$ and $k_{i,r}$ with a certain resolution $\Delta\tau = \tau_{\max}/Q$ and $\Delta\omega = \omega_{\max}/K$. The search spaces for the MF receiver then become $\mathcal{Q} = \{0, 1, \cdots, Q-1\}$ and $\mathcal{K} = \{-K, \cdots, K\}$, which is the discrete counterpart of the continuous search space $\mathcal{T} \times \mathcal{F}$.

The MF receiver is a popular choice for multiuser acquisition [3], for example, in GPS receivers [27] or CDMA receivers [25]. Its comparison with the C-SA acquisition scheme using the CMS architecture we propose in this paper is particularly insightful because, although the MF front-end requires a large filterbank, the post-processing of its outputs is very simple. One could certainly perform more complex post-processing to enhance its performance. For example, the Orthogonal Matching Pursuit (OMP) algorithm in our C-SA scheme can be applied on the MF outputs for this purpose. However, in that case, as illustrated in Section VIII, the resulting scheme will have much higher storage cost and computational complexity requirements. More importantly, the OMP technique can be directly applied to the Nyquist samples, as done in SA methods, making the MF stage superfluous[1].

The MF obtains the decision statistics by passing $x(t)$ through a bank of $P = I|\mathcal{K}|$ MF templates, and sampling the outputs every $\Delta\tau$. To be consistent with the sequential structure in (5), the MF shifts its templates every $D = NT_s$, and samples the outputs every $\Delta\tau = T_s \leq 1/f_{\text{NYQ}}$. The output of the pth MF with $p = (k + K)I + i$ for $k \in \mathcal{K}$

[1]Strictly speaking, the OMP technique performs a MF stage in its first iteration. The subsequent iterations can be viewed as applying successive cancellation.

and $i = 1, \cdots, I$ is obtained as

$$c_{i,k}[w] = \left\langle x(t), \phi_i(t - wT_s)e^{\mathrm{i}k\Delta\omega(t-wT_s)} \right\rangle. \quad (16)$$

Oftentimes, the filtering process is implemented in the digital domain using the samples $\mathbf{c}_{\text{DS}}[n]$ in (8). For consistency, we proceed with the description in the analog domain. At the nth shift, the samples used for detections can be stacked into an $I|\mathcal{K}| \times |\mathcal{Q}|$ exhaustive MF output array

$$\mathbf{C}_{\text{MF}}[n] = \begin{bmatrix} \vdots & \cdots & \vdots \\ c_{1,k}[nN] & \cdots & c_{1,k}[nN+Q-1] \\ \vdots & \cdots & \vdots \\ c_{I,k}[nN] & \cdots & c_{I,k}[nN+Q-1] \\ \vdots & \cdots & \vdots \end{bmatrix}. \quad (17)$$

Then the MF receiver uses $\mathbf{C}_{\text{MF}}[n]$ as test statistics and performs the test on each user as follows

$$|c_{i,k}[nN+q]| \geq \rho_i, \ i = 1, \cdots, I, \ k \in \mathcal{K}, \ q \in \mathcal{Q},$$

where ρ_i is the chosen detection threshold for each user.

Denote the active user set at the nth shift as

$$\widehat{\mathcal{I}} = \{i : |c_{i,k}[nN+q]| \geq \rho_i, \ \forall i,k,q\}$$

and the shift $N_\eta \triangleq \min\left\{\arg\limits_n \ |c_{i,k}[nN+q]| \geq \rho_i, \forall i,k,q\right\}$. The MLR shift is then obtained as

$$\ell_\star \triangleq \arg\max_n \left[\max_{i,k,q} \ |c_{i,k}[nN+q]|\right], \ n = N_\eta, \cdots, N_\eta + N_0.$$

Given the multipath order R, the delay-Doppler pairs at the nth shift are found as the R strongest outputs $c_{i,k}[\ell_\star N+q]$'s for all detected users $i \in \widehat{\mathcal{I}}$ over the MF search space $k \in \mathcal{K}$ and $q \in \mathcal{Q}$. For convenience, we denote the strongest path by $(k_{i,1}, q_{i,1})$ for the ith user $i \in \widehat{\mathcal{I}}$ and order the samples by magnitudes

$$\begin{aligned} |c_{i,k_{i,1}}[\ell_\star N + q_{i,1}]| > |c_{i,k_{i,2}}[\ell_\star N + q_{i,2}]| > \cdots \quad (18) \\ > |c_{i,k_{i,R}}[\ell_\star N + q_{i,R}]| > \cdots \end{aligned}$$

for each user $i \in \widehat{\mathcal{I}}$. The delay-Doppler pairs are identified as the set

$$\mathcal{M}_i \triangleq \{(k_{i,r}, q_{i,r}) : r = 1, \cdots, R\}, \quad (19)$$

which give the following link parameters

$$\hat{\tau}_{i,r} = q_{i,r}\Delta\tau, \ \hat{\omega}_{i,r} = k_{i,r}\Delta\omega, \quad (k_{i,r}, q_{i,r}) \in \mathcal{M}_i.$$

Although the MF approach shows an advantage in its post-processing and implementation, it has a few drawbacks:

i) the size of the MF filterbank scales with the number of users I and the parameter set $|\mathcal{K}|$;
ii) digital implementation requires high rate processing, increasing storage and pipelining[2] cost;
iii) the MF samples $c_{i,k}[\cdot]$ contain the interference from different users and multipath components;

[2]Pipelining refers to timely processing of the samples that stream into the system per unit of time.

During the link acquisition phase, the effect of interference (iii) is mitigated by using wideband pulses $g(t)$. During the data detection phase, multipath and multiuser interferences are dealt with using a RAKE type receiver and the interference is tackled either by using linear multiuser receivers or, in some cases, using successive interference cancelation (SIC) or even maximum likelihood multiuser detection [3]. Typically, the complexities of these schemes for data detection grow rapidly with respect to the size of the MF filterbank (i) and the sampling rate (ii). Since this phase is conducted after the link acquisition, the uncertainties about the set of active users, their delays and Doppler frequencies have already been resolved and therefore, these tasks become more manageable.

C. Sparsity Aware (SA) Approach

Instead of simply observing and ranking the MF outputs, many recent works have proposed the idea of *compressed sensing* or *sparse recovery* to solve estimation and detection problems. For the purpose of *user identification and parameter estimation*, one approach is to approximate (9) by a sparse model with a dictionary constructed from the ensemble of possible templates $\phi_i(t)$ [1], [4] and/or discretized delays $\boldsymbol{\tau}$ [2] (similar to the MF templates), where the joint recovery of active users and unknown parameters is relaxed as a *sparse estimation* problem. These sparse methods, which we call the *Direct* Sparsity-Aware (D-SA) scheme, usually assume the MLR shift $\ell_\star$ to be known and require Nyquist rate samples. On the other hand, aiming at *signal presence detection* rather than identifying the active users and recovering the parameters, [12]–[16] reduce the number of samples required for the test by using a linear *compressor* on the block of given discrete observations, which we call the *Compressive* Sparsity-Aware (C-SA) scheme. In terms of the acquisition front-end, the D-SA scheme is a special case of C-SA scheme with a compressor that is an identity matrix.

In this paper, the proposed CMS architecture with SR-LRT bears resemblances of certain features of the C-SA scheme because of the compressive samples obtained in the front-end. Thus, to avoid any confusion, we also refer to the SR-LRT based on the CMS architecture in this paper as the C-SA scheme. However, a distinctive difference between the C-SA scheme in this paper and those in [12]–[16] is that the C-SA scheme proposed in this paper unifies the sequential signal detection, identification of active users and estimation of parameters by using the compressive samples obtained from a flexible multi-rate A/D architecture, while the C-SA scheme in [12]–[16] directly starts from an abstract discrete model that is already sampled. Last but not least, the sampling kernels in the proposed CMS architecture are further optimized with respect to the estimation and detection performance.

IV. Compressive Sequential Link Acquisition

A. Compressive Multichannel Sampling (CMS)

We propose to use the A/D front-end in Fig. 1, typical in FRI sampling [17], [19], [20], [22], [23] for this work, which samples the signal every $t = nD$ by a P-channel filterbank

$$c_p[n] \triangleq \langle x(t), \psi_p(t - nD)\rangle, \quad p = 1, \cdots, P. \quad (20)$$

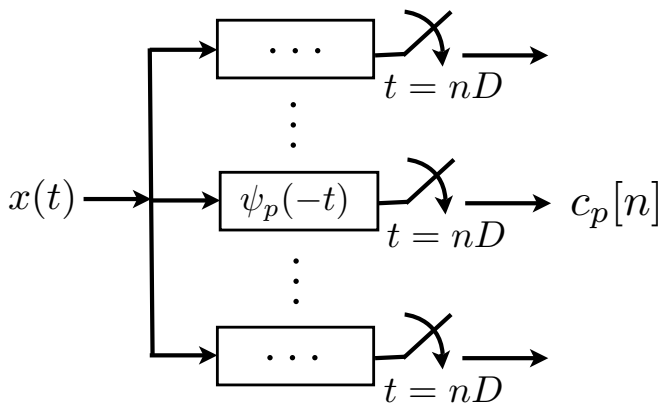


Fig. 1. Samples obtained from compressive acquisition.

We call this architecture the Compressive Multichannel Sampling (CMS) scheme.

Note that (20) can also be implemented in the digital domain by performing linear projections of the discrete signal $\mathbf{c}_{\text{DS}}[n]$ in (9). This means that the CMS architecture becomes part of the post-processing of the Nyquist samples of $x(t)$, which lowers the storage and computation requirements as illustrated in Section VIII. Similar derivations can be done in discrete time, but the advantage of using the analog description is that we do not necessarily have to target bandlimited signals. In the FRI literature [17]–[20], [22], [23], in the absence of noise, the sampling rate required for the unique reconstruction of the signal in (5) is the *number of degrees of freedom* of the signal $x(t)$ per shift D, equal to the number of unknowns $\{\tau_{i,r}, \omega_{i,r}, h_{i,r}\}_{i\in\mathcal{I}, r=1,\cdots,R}$. This amounts to $P_{\min} = 3|\mathcal{I}|R$, which can be much less than what is needed in the MF approach $P_{\min} \ll P_{\text{MF}} = I|\mathcal{K}|$ when the number of active users is not large $|\mathcal{I}| \ll I$, or the number of multipaths R is much less than the dimension of the search space for Doppler $|\mathcal{K}|$. However, since the estimation and detection are performed in the presence of noise, the number of P needs to be increased in general to enhance the sensitivity of the receiver. This gives the option of trading off accuracy with storage cost and computational complexity, by adjusting the number of samples P to process between $P_{\min} \le P \le P_{\text{MF}}$.

Note that for different schemes, we need further processing to produce final decisions. This last step is different for different receivers. For example, the MF scheme has very simple post-processing at the cost of handling exhaustive MF samples, while the C-SA scheme can tune the number of measurements to handle less data by spending a higher premium for sparsity recovery algorithms. Therefore, we discuss this trade-off specifically in detail in Section VIII, with the MF being a benchmark for our comparison.

B. CMS Observation Model

Similar to [1], [2], we follow the analog description of (5) but discretize the parameters as in the MF approach. The analog domain derivation is mostly inspired by the FRI literature, but it is not tied to wanting to estimate continuous parameters here in this paper. For notational convenience, we introduce the triple-index coefficient

$$\alpha_{i,k,q} = \sum_{j\in\mathcal{I}} \sum_{r=1}^{R} h_{j,r}\delta[i-j]\delta[k-k_{j,r}]\delta[q-q_{j,r}], \qquad (21)$$

for $k \in \mathcal{K}, q \in \mathcal{Q}$ as an indicator of whether the ith user is transmitting and whether there exists a link at a certain delay $\tau = q\Delta\tau$ with a certain carrier offset $\omega = k\Delta\omega$ in the window. Note that $\alpha_{i,k,q} = 0$ except when $k = k_{i,r}$ and $q = q_{i,r}$ for $i \in \mathcal{I}$. Denoting each *MF template* by

$$\phi_{i,k,q}(t) \triangleq \phi_i(t - q\Delta\tau)e^{\mathrm{i}k\Delta\omega t}, \qquad (22)$$

the signal in (5) can be approximately expressed as

$$x(t) = \sum_{i=1}^{I} \sum_{k\in\mathcal{K}} \sum_{q\in\mathcal{Q}} \alpha_{i,k,q} e^{\mathrm{i}k\Delta\omega\ell D}\ \phi_{i,k,q}(t - \ell D) + v(t). \qquad (23)$$

Clearly, $x(t)$ has at most $|\mathcal{I}|R$ active components due to the sparsity of $\alpha_{i,k,q}$. To facilitate notations in our derivations, we introduce the triplet index (i,k,q) and define the length-$I|\mathcal{K}||\mathcal{Q}|$ *link vector* $\boldsymbol{\alpha}[\ell]$ at the ℓth shift as

$$\Big[\boldsymbol{\alpha}[\ell]\Big]_{(i,k,q)} \triangleq \Big[\boldsymbol{\alpha}[\ell]\Big]_{(i-1)|\mathcal{K}||\mathcal{Q}|+(k-1)|\mathcal{Q}|+q} = \alpha_{i,k,q}. \qquad (24)$$

We define the associated *delay-Doppler set* for the ith user at the ℓth shift as

$$\mathcal{A}_i^{(\ell)} \triangleq \{(k,q) : |\alpha_{i,k,q}| > 0, k \in \mathcal{K}, q \in \mathcal{Q}\}, \quad i = 1, \cdots, I, \qquad (25)$$

from which we extract the time delays $\tau_{i,r} = q_{i,r}\Delta\tau$ and Doppler frequencies $\omega_{i,r} = k_{i,r}\Delta\omega$ for the active users $i \in \mathcal{I}$ if $\mathcal{A}_i^{(\ell)} \neq \varnothing$.

In this work, we consider the use of sampling kernels $\psi_p(t)$ that are linear combinations of all the MF templates [23]. The following theorem specifies the observed samples from the CMS architecture in Fig. 1 in relation to the link vector $\boldsymbol{\alpha}[\ell]$.

Theorem 1. *Suppose that we choose sampling kernels $\{\psi_p(t)\}_{p=1}^{P}$ as linear combinations of the MF templates*

$$\psi_p(t) = \sum_{i=1}^{I} \sum_{k\in\mathcal{K}} \sum_{q\in\mathcal{Q}} b_{p,(i,k,q)}\phi_{i,k,q}(t), \quad p = 1, \cdots, P. \qquad (26)$$

The length-P sample vector $\mathbf{c}[n] \triangleq [c_1[n], \cdots, c_P[n]]^T$ taken at shift $t = nD$, can then be expressed as

$$\mathbf{c}[n] = \mathbf{B}\mathbf{M}_{\phi\phi}[n-\ell]\boldsymbol{\Gamma}[\ell]\boldsymbol{\alpha}[\ell] + \boldsymbol{\nu}[n], \qquad (27)$$

where $\boldsymbol{\alpha}[\ell]$ is the link vector at the ℓth shift and

1) *$\mathbf{B}$ is a $P \times I|\mathcal{K}||\mathcal{Q}|$ matrix with $[\mathbf{B}]_{p,(i,k,q)} \triangleq b_{p,(i,k,q)}$; $\mathbf{M}_{\phi\phi}[n-\ell]$ is an $I|\mathcal{K}||\mathcal{Q}| \times I|\mathcal{K}||\mathcal{Q}|$ matrix with*

$$\Big[\mathbf{M}_{\phi\phi}[n-\ell]\Big]_{(i',k',q'),(i,k,q)} \triangleq R_{\phi_{i',k',q'}\phi_{i,k,q}}[(n-\ell)D],$$

where

$$R_{\phi_{i',k',q'}\phi_{i,k,q}}[\Delta t] = e^{\mathrm{i}k\Delta\omega(\Delta t - q'\Delta\tau)} R_{\phi_{i'},\phi_i}^{(k-k')}\left[(q'-q)\Delta\tau + \Delta t\right],$$

and $R^{(k-k')}_{\phi_{i'},\phi_i}(\cdot)$ *is the ambiguity function*

$$R^{(k-k')}_{\phi_{i'},\phi_i}(\cdot) = \int \phi^*_{i'}(t)\phi_i(t-\cdot)e^{\mathrm{i}(k-k')\Delta\omega t}\mathrm{d}t. \quad (28)$$

2) $\mathbf{\Gamma}[\ell] \triangleq \mathbf{I}_I \otimes \mathbf{E}[\ell] \otimes \mathbf{I}_{|\mathcal{Q}|}$ *with*

$$\mathbf{E}[\ell] \triangleq \mathrm{diag}[e^{\mathrm{i}K\Delta\omega\ell D}, \cdots, e^{-\mathrm{i}K\Delta\omega\ell D}]; \quad (29)$$

3) $\boldsymbol{\nu}[n] \triangleq [\nu_1[n], \cdots, \nu_P[n]]^T$ *is the filtered Gaussian noise vector with zero mean and covariance*

$$\mathbf{R}_{vv} = \sigma^2\mathbf{R}_{\psi\psi} \text{ with } \mathbf{R}_{\psi\psi} = \mathbf{B}\mathbf{M}_{\phi\phi}[0]\mathbf{B}^H. \quad (30)$$

Proof: See Appendix A. ∎

The freedom in choosing $\mathbf{B}$ allows us to optimize link acquisition performance. Before discussing the details of optimization in Section VI, we further simplify the model in Theorem 1.

C. CMS Sequential Acquisition Model

Theorem 1 describes the general model of the samples $\mathbf{c}[n]$ obtained in the nth shift with respect to the link vector $\boldsymbol{\alpha}[\ell]$. However, the exact shift ℓ is unknown to the receiver. As mentioned earlier, determining the exact shift is not necessary to recover the link parameters, as long as the shift is properly aligned with the signal and produces a positive detection maximizing the likelihood ratio. In the following, we transform the observation model $\mathbf{c}[n]$ in Theorem 1 to an equivalent model. The equivalent model is stated with respect to a modified link vector $\boldsymbol{\alpha}[n]$ at the nth shift, which contains entries that are shifted with the relative placement of $(n-\ell)$ in relation to $\boldsymbol{\alpha}[\ell]$. The reason for this is that we can use a time-invariant system matrix instead of a time-variant one $\mathbf{M}_{\phi\phi}[n-\ell]$ for the purpose of sequential detection.

Theorem 2. *Let* $D = N\Delta\tau$ *for some integer* $N \in \mathbb{Z}$. *The outputs* $\mathbf{c}[n]$ *of the compressive samplers can be re-written as*

$$\mathbf{c}[n] = \mathbf{B}\mathbf{M}\mathbf{\Gamma}[n]\boldsymbol{\alpha}[n] + \boldsymbol{\nu}[n], \quad (31)$$

where $\mathbf{M} \triangleq \mathbf{M}_{\phi\phi}[0]$, *and* $\boldsymbol{\alpha}[n]$ *is the link vector at the* n*th shift*

$$\Big[\boldsymbol{\alpha}[n]\Big]_{(i,k,q)} \triangleq \alpha_{i,k,q+(n-\ell)N}. \quad (32)$$

Proof: See Appendix B. ∎

Corollary 1. *Let the delay-Doppler sets at the* n*th shift be*

$$\mathcal{A}^{(n)}_i \triangleq \left\{(k,q) : \left|\Big[\boldsymbol{\alpha}[n]\Big]_{(i,k,q)}\right| \neq 0, k \in \mathcal{K}, q \in \mathcal{Q}\right\} \quad (33)$$

for $i = 1, \cdots, I$. *Then for any* $(k,q) \in \mathcal{A}^{(\ell)}_i$ *at the* ℓ*th shift* (25), *we have* $(k, q-(n-\ell)N) \in \mathcal{A}^{(n)}_i$ *at the* n*th shift.*

Using the modified sets $\mathcal{A}^{(n)}_i$, the number of delay-Doppler pairs included at the nth shift equals $\sum_{i=1}^{I}\left|\mathcal{A}^{(n)}_i\right|$. It is obvious that $\left|\mathcal{A}^{(n)}_i\right| \leq \left|\mathcal{A}^{(\ell)}_i\right|$ for any i. At the nth shift, if a positive detection is declared and $\left|\mathcal{A}^{(n)}_i\right| = \left|\mathcal{A}^{(\ell)}_i\right|$ for all i, then the modified link vector $\boldsymbol{\alpha}[n]$ at the nth shift carries equivalent link information as the link vector $\boldsymbol{\alpha}[\ell]$ at the ℓth shift. Therefore, we use the model in Theorem 2 for our design and re-state the goal of *link acquisition* as

1) locating the MLR shift $\ell_\star$;
2) identifying the set of active users $\widehat{\mathcal{I}}$ indicated by the *delay-Doppler set* $\mathcal{A}^{(\ell_\star)}_i \neq \varnothing$;
3) resolving the delay-Doppler pairs in the $\ell_\star$th window $\mathcal{A}^{(\ell_\star)}_i \subseteq \mathcal{K}\times\mathcal{Q}$ for $i \in \widehat{\mathcal{I}}$.

For better representation and comparison of the individual support set $\mathcal{A}^{(n)}_i$ in relation to the original support set $\mathcal{A}^{(\ell)}_i$, we introduce the full *user-delay-Doppler sets* for the link vectors $\boldsymbol{\alpha}[n]$ and $\boldsymbol{\alpha}[\ell]$ respectively

$$\mathcal{A}_n \triangleq \left\{(i,k,q) : \ (k,q) \in \mathcal{A}^{(n)}_i, i \in \mathcal{I}\right\}, \quad (34)$$

$$\mathcal{A}_\ell \triangleq \left\{(i,k,q) : \ (k,q) \in \mathcal{A}^{(\ell)}_i, i \in \mathcal{I}\right\}. \quad (35)$$

In the following sections, we express the link vector explicitly with respect to the full *user-delay-Doppler set* $\mathcal{A}_n$ and combine the phase rotation matrix $\mathbf{\Gamma}[n]$ at the nth shift as

$$\boldsymbol{\beta}_{\mathcal{A}_n} \triangleq \mathbf{\Gamma}[n]\boldsymbol{\alpha}[n]. \quad (36)$$

We call $\boldsymbol{\beta}_{\mathcal{A}_n} = [\cdots, \beta_{i,k,q}, \cdots]^T$ the *modified link vector* and note that it is also a $|\mathcal{I}|R$-sparse vector.

V. Compressive Sequential Link Acquisition with Sparsity Regularization

We now develop an SR-LRT detection algorithm that tackles the link acquisition problem exploiting the compressive observation model given in Theorem 2. Link acquisition attempts to discriminate the true pattern $\mathcal{A}_n$ against all possible patterns $\mathcal{S}_n \neq \mathcal{A}_n$ at every shift $t = nD$ as a hypothesis testing problem

$$\mathcal{H}_{\mathcal{S}_n}: \quad \mathbf{c}[n] = \mathbf{B}\mathbf{M}\boldsymbol{\beta}_{\mathcal{S}_n} + \mathbf{v}[n] \quad (37)$$

over all possible $\mathcal{S}_n$ at every shift $t = nD$. Note that the signal presence detection is also incorporated in this test by choosing $\mathcal{S}_n = \varnothing$ to be the null hypothesis. Given a specific set $\mathcal{S}_n$ for each possible $\boldsymbol{\beta}_{\mathcal{S}_n}$, the amplitudes of $\boldsymbol{\beta}_{\mathcal{S}_n}$ and the noise variance σ^2 are unknown and treated as nuisance parameters. The link acquisition is thus to detect the full *user-delay-Doppler set* $\mathcal{S}_n$ for all possible $\mathcal{H}_{\mathcal{S}_n}$ with $\boldsymbol{\beta}_{\mathcal{S}_n}$ and the noise level σ^2 being nuisance in each shift n. Following the GLRT rationale, the test consists in finding the set $\mathcal{S}_n$ maximizing

$$\mathbb{P}\left(\mathcal{H}_{\mathcal{S}_n}|\boldsymbol{\beta}_{\mathcal{S}_n}, \sigma^2\right) \quad (38)$$

$$= \frac{1}{\pi^P\sigma^{2P}|\mathbf{R}_{\psi\psi}|}\exp\left(-\frac{\left\|\mathbf{c}[n] - \mathbf{B}\mathbf{M}\boldsymbol{\beta}_{\mathcal{S}_n}\right\|^2_{\mathbf{R}^{-1}_{\psi\psi}}}{\sigma^2}\right) \quad (39)$$

in the presence of unknown parameters $\boldsymbol{\beta}_{\mathcal{S}_n}$ and σ^2.

Note that when $\mathbf{B} = \mathbf{I}$ is identity, the samples $\mathbf{c}[n]$ are equivalent to the outputs of the MF approach. This implies that the samples $\mathbf{c}[n]$ obtained in the CMS architecture, using only P sampling kernels in (26), are equivalent to a linearly compressed version of the exhaustive MF output, even though they are obtained directly from the A/D architecture instead of using an exhaustive MF filterbank followed by a linear

compressor $\mathbf{B}$ as in [14], [15], which would be much more complex. On the other hand, the difference in post-processing between the MF and CMS architecture is that MF allows to simply pick the hypothesis corresponding to the largest magnitude in the output $\mathbf{c}[n]$ as the detection result, while using the compressive samples $\mathbf{c}[n]$ one necessarily needs a more sophisticated detection scheme.

A. Sequential Estimation for Link Acquisition

The GLRT involved in the link acquisition requires estimating $\boldsymbol{\beta}_{\mathcal{S}_n}$ and σ^2 for every possible $\mathcal{S}_n$ at every shift $t = nD$. For every hypothesis $\mathcal{S}_n \neq \varnothing$, the estimate of $\boldsymbol{\beta}_{\mathcal{S}_n}$ under colored Gaussian noise $\boldsymbol{\nu}[n]$ with covariance $\mathbf{R}_{vv} = \sigma^2 \mathbf{R}_{\psi\psi}$ is then obtained as

$$\widehat{\boldsymbol{\beta}}_{\mathcal{S}_n} \triangleq \arg\min_{\boldsymbol{\beta}_{\mathcal{S}_n}} \left\| \mathbf{c}[n] - \mathbf{B}\mathbf{M}\boldsymbol{\beta}_{\mathcal{S}_n} \right\|^2_{\mathbf{R}^{-1}_{\psi\psi}}. \tag{40}$$

The "hat" notation $\widehat{(\cdot)}$ on the vector $\boldsymbol{\beta}_{\mathcal{S}_n}$ refers to the estimates of the amplitudes on the support $\mathcal{S}_n$. The total number of such estimates scales with the number of hypothesis which in this case is $2^{I|\mathcal{K}||\mathcal{Q}|}$, resulting in an NP-hard combinatorial estimation problem. Instead, we obtain the estimates using a sparse approach similar to [1], [2] for the GLRT. Specifically, we solve the combinatorial problem in a "soft" fashion at every shift $t = nD$ similar to [2]

$$\widehat{\boldsymbol{\beta}} \triangleq \arg\min_{\boldsymbol{\beta}} \ \|\mathbf{c}[n] - \mathbf{B}\mathbf{M}\boldsymbol{\beta}\|^2_{\mathbf{R}^{-1}_{\psi\psi}} + \lambda \cdot f(\boldsymbol{\beta}), \tag{41}$$

where λ is some regularization parameter and $f(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_0$ or $f(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$ are the sparsity regularization constraint. If the $\|\cdot\|_0$ constraint is imposed, the problem is approximately solved via greedy methods such as orthogonal matching pursuit (OMP) [28]. When $\|\cdot\|_1$ norm is used, this problem can be solved via convex programs [29]. Generally speaking, as discussed in [2] and [29], the required number of samples P for sparse recovery in the noiseless case scales logarithmically with the dimension of the length-$I|\mathcal{K}||\mathcal{Q}|$ link vector $P \propto |\mathcal{I}|R\log I|\mathcal{K}||\mathcal{Q}|$, which increases if discretizations are made finer.

From the solution of (41), we extract the full *user-delay-Doppler set* $\widehat{\mathcal{A}}_n$ from the soft estimate $\widehat{\boldsymbol{\beta}}$, which in turns gives the estimated user set $\widehat{\mathcal{I}}$ and the estimated *delay-Doppler set* $\widehat{\mathcal{A}}_i^{(n)}$ for each user $i \in \widehat{\mathcal{I}}$

$$\mathcal{E}\left(\widehat{\boldsymbol{\beta}}\right) = \widehat{\mathcal{A}}_n \quad \Longrightarrow \quad \widehat{\mathcal{I}} \text{ and } \left\{\widehat{\mathcal{A}}_i^{(n)}\right\}_{i\in\widehat{\mathcal{I}}}. \tag{42}$$

Here $\mathcal{E}(\cdot)$ is the extraction mapping from the soft estimate to the estimated *user-delay-Doppler* set. The extraction method is explained in Section V-B.

With the estimated set of active users $\widehat{\mathcal{I}}$ and individual *delay-Doppler set* $\widehat{\mathcal{A}}_i^{(n)}$, we have the truncated estimate of the link vector $\widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{A}}_n}$ and the estimated noise variance

$$\widehat{\sigma}^2_{\widehat{\mathcal{A}}_n} = \left\| \mathbf{c}[n] - \mathbf{B}\mathbf{M}\widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{A}}_n} \right\|^2_{\mathbf{R}^{-1}_{\psi\psi}}/P. \tag{43}$$

Since the formulation in (41) is no longer maximum likelihood due to the sparsity regularization, we call it the Sparsity-Regularized Likelihood Ratio Test (SR-LRT).

B. User-Delay-Doppler Set Extraction $\mathcal{E}\left(\widehat{\boldsymbol{\beta}}\right)$

Given the soft estimate $\widehat{\boldsymbol{\beta}}$ in (41) at every shift $t = nD$, the estimated user-delay-Doppler set $\widehat{\mathcal{A}}_n$ is extracted depending on the application scenarios below.

1) Unknown, random number of active users $\mathcal{I}$: In random access communications the receiver has no knowledge of who is active, nor any expectation on the number of components it is likely to detect. Using this soft estimate $\widehat{\boldsymbol{\beta}}$, we identify the active users as

$$\widehat{\mathcal{I}} \triangleq \left\{ i : \max_{k,q} \ \left|\widehat{\beta}_{i,k,q}\right|^2 \geq \rho_i, k \in \mathcal{K}, q \in \mathcal{Q} \right\} \tag{44}$$

where ρ_i is a chosen threshold for that specific user to be considered present, usually set as a fraction of the magnitude of the amplitudes in $\widehat{\boldsymbol{\beta}}$. Then for each detected active user $i \in \widehat{\mathcal{I}}$, we take R strongest paths in $\widehat{\beta}_{i,k,q}$ with respect to $k \in \mathcal{K}$ and $\mathcal{Q}$ to be the active set $\widehat{\mathcal{A}}_i^{(n)}$ for each user $i \in \widehat{\mathcal{I}}$.

2) Partial knowledge on active users $\mathcal{I}$: This scenario corresponds to environments where all users are active, however only a certain subset is likely to be detectable by the receiver. GPS receivers are an example. Specifically, there are a total of $I = 24$ quasi-stationary GPS satellites moving around the earth and the active satellites in the field-of-view of a specific GPS receiver are unknown. However, the GPS receiver is informed that at any point in space there should be $|\mathcal{I}| = 4$ strongest signals from satellites, and it attempts to find such signals, along with their delay-Doppler parameters for triangularization. In this case a positive detection corresponds to having at least four components detected and we can interpret this case as fixing $|\mathcal{I}|$ for the receiver detection. In general, we identify the users $i \in \widehat{\mathcal{I}}$ as those with the $|\mathcal{I}|$ strongest amplitudes $|\widehat{\beta}_{i,k,q}|$ with respect to $i = 1, \cdots, I$ in $\widehat{\boldsymbol{\beta}}$. Then we take R strongest paths in $|\widehat{\beta}_{i,k,q}|$ with respect to k and q to be the active set $\widehat{\mathcal{A}}_i^{(n)}$ for each user $i \in \widehat{\mathcal{I}}$.

3) Known active users $\mathcal{I}$: This scenario includes multi-antenna and cooperative transmission systems, where the receiver is aware of the active sources, i.e., $\mathcal{I}$ is known. This case is trivial because we do not need to identify the active users. The active set $\widehat{\mathcal{A}}_i^{(n)}$ for each user $i \in \mathcal{I}$ is formed by picking the R strongest components in $|\widehat{\beta}_{i,k,q}|$ with respect to $k \in \mathcal{K}$ and $\mathcal{Q}$.

C. Sequential Detection for Link Acquisition

Substituting $\widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{A}}_n}$ and $\widehat{\sigma}^2_{\widehat{\mathcal{A}}_n}$ back to (38), the generalized likelihood ratio can be computed as

$$\eta_{\text{C-SA}}(n) \triangleq \frac{\mathbb{P}\left(\mathcal{H}_{\widehat{\mathcal{A}}_n} | \widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{A}}_n}, \widehat{\sigma}^2_{\widehat{\mathcal{A}}_n}\right)}{\mathbb{P}\left(\mathcal{H}_{\varnothing} | \widehat{\sigma}^2_{\varnothing}\right)} \tag{45}$$

$$= \frac{\|\mathbf{c}[n]\|^{2P}_{\mathbf{R}^{-1}_{\psi\psi}}}{\left\|\mathbf{c}[n] - \mathbf{B}\mathbf{M}\widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{A}}_n}\right\|^{2P}_{\mathbf{R}^{-1}_{\psi\psi}}} > \eta_0, \tag{46}$$

which indicates the presence of the signal if $\eta_{\text{C-SA}}(n) > \eta_0 \geq 1$ so that the receiver knows that certain signal components are captured in the observation. Denote the first window that

passes the above test as $N_\eta \triangleq \min\left\{\arg\limits_n \ \eta_{\text{C-SA}}(n) \geq \eta_0\right\}$. As mentioned in (15), the MLR window is located as the window that maximizes the likelihood ratio

$$\ell_\star = \arg\max_n \ \ \eta_{\text{C-SA}}(n), \quad n = N_\eta, \cdots, N_\eta + N_0. \tag{47}$$

Accordingly, from the link vector $\widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{A}}_{\ell_\star}}$ in the $\ell_\star$th window, we can extract the delay-Doppler pairs

$$\widehat{\tau}_{i,r} = q_{i,r}\Delta\tau, \ \widehat{\omega}_{i,r} = k_{i,r}\Delta\omega, \ (k,q) \in \widehat{\mathcal{A}}_i^{(\ell_\star)}, \ i \in \widehat{\mathcal{I}}. \tag{48}$$

VI. Optimization of Compressive Samplers

The link acquisition performance depends on the ability of the SR-LRT to differentiate between different hypotheses $\mathcal{H}_{\mathcal{S}_n}$. In this section, we seek a criterion to optimize the sampling kernels $\{\psi_p(t)\}_{p=1}^P$ by designing the matrix $\mathbf{B}$. The metric we maximize is the weighted average of the Kullback-Leibler (KL) distances between any $\mathcal{H}_{\mathcal{S}_n}$ in (37). Since every possible pattern for $\mathcal{S}_n$ is independent of n, here we omit the subscript for convenience. In choosing the KL distance we are motivated by the Chernoff-Stein's lemma [30], whose statement indicates that the probability of confusing $\mathcal{H}_{\mathcal{S}}$ and $\mathcal{H}_{\mathcal{S}'}$ decreases exponentially with the pair-wise KL distance between them. As we point out later in this section, if the noise is Gaussian and the weights are chosen appropriately, then the weighted average KL distance of all the *pair-wise KL* distances has the same expression as the Chernoff information under the Bayesian detection framework, implying that the average KL distance is an effective measure in evaluating detection performance. Being consistent with our system model and detection formulation, we proceed with our analysis using the average KL distance with some pre-defined weights.

Introducing $\mathcal{G}(\mathbf{B}) \triangleq \mathbf{M}^H\mathbf{B}^H\left(\mathbf{B}\mathbf{M}\mathbf{B}^H\right)^{-1}\mathbf{B}\mathbf{M}$, the *pair-wise KL* distance between any $\mathcal{H}_{\mathcal{S}}$ and $\mathcal{H}_{\mathcal{S}'}$ is given by [31]

$$\mathbb{D}\left(\mathcal{H}_{\mathcal{S}} \| \mathcal{H}_{\mathcal{S}'}\right) = \frac{\left(\boldsymbol{\beta}_{\mathcal{S}} - \boldsymbol{\beta}_{\mathcal{S}'}\right)^H \mathcal{G}(\mathbf{B}) \left(\boldsymbol{\beta}_{\mathcal{S}} - \boldsymbol{\beta}_{\mathcal{S}'}\right)}{\sigma^2}. \tag{49}$$

If the pair-wise KL distance is zero, then the two hypotheses $\mathcal{H}_{\mathcal{S}}$ and $\mathcal{H}_{\mathcal{S}'}$ are indistinguishable for that particular pairs of $\mathcal{S}$ and $\mathcal{S}'$. A non-zero pair-wise KL distance between arbitrary pair of $\mathcal{S}$ and $\mathcal{S}'$ with $|\mathcal{S}|, |\mathcal{S}'| \leq s$ requires $\text{spark}\left[\mathcal{G}(\mathbf{B})\right] \geq 2s$, where $\text{spark}[\cdot]$ is the kruskal rank of a matrix. We note that the average KL distance metric we are about to define, does not automatically ensure that $\mathbb{D}\left(\mathcal{H}_{\mathcal{S}} \| \mathcal{H}_{\mathcal{S}'}\right) > 0$ for all $\mathcal{S} \neq \mathcal{S}'$.

To define the average KL distance, we associate each distinct pair of supports $\mathcal{S}$ and $\mathcal{S}'$ with the weight $\gamma_{\mathcal{S},\mathcal{S}'}$. Furthermore, we associate the nuisance amplitudes in $\boldsymbol{\beta}_{\mathcal{S}}$ a multidimensional continuous weighting function $P(\boldsymbol{\beta}_{\mathcal{S}})$ for any $\mathcal{S}$. Under these assumptions, the weighted average of all *pair-wise KL* distances is defined as

$$\overline{\mathbb{D}} = \sum_{\mathcal{S}} \sum_{\mathcal{S}' \neq \mathcal{S}} \gamma_{\mathcal{S},\mathcal{S}'} \iint P(\boldsymbol{\beta}_{\mathcal{S}}) P(\boldsymbol{\beta}_{\mathcal{S}'}) \mathbb{D}\left(\mathcal{H}_{\mathcal{S}} \| \mathcal{H}_{\mathcal{S}'}\right) \mathrm{d}\boldsymbol{\beta}_{\mathcal{S}} \mathrm{d}\boldsymbol{\beta}_{\mathcal{S}'}. \tag{50}$$

Proposition 1. *Given a set of normalized constant weights $\gamma_{\mathcal{S},\mathcal{S}'}$ for every distinct pair $\mathcal{S}, \mathcal{S}'$, and a continuous weighting function $P(\boldsymbol{\beta}_{\mathcal{S}}) = \prod_{(i,k,q)\in\mathcal{S}} P(\beta_{i,k,q})$ over the nuisance amplitudes with $\int \boldsymbol{\beta}_{\mathcal{S}} P(\boldsymbol{\beta}_{\mathcal{S}}) \mathrm{d}\boldsymbol{\beta}_{\mathcal{S}} = \mathbf{0}$ and $\int |\beta_{i,k,q}|^2 P(\beta_{i,k,q}) \mathrm{d}\beta_{i,k,q} =$ constant, the average KL distance $\overline{\mathbb{D}}$ is equal to*

$$\overline{\mathbb{D}} = \frac{1}{\sigma^2} \text{Tr}\left[\mathbf{M}^H\mathbf{B}^H\left(\mathbf{B}\mathbf{M}\mathbf{B}^H\right)^{-1}\mathbf{B}\mathbf{M}\right]. \tag{51}$$

Proof: See Appendix C. ∎

The way we choose the weights $\gamma_{\mathcal{S},\mathcal{S}'}$ and weighting functions $P(\boldsymbol{\beta}_{\mathcal{S}})$ is equivalent to assuming a uniform distribution on the users, delays and Dopplers together with i.i.d. Gaussian priors on the amplitudes in $\boldsymbol{\beta}_{\mathcal{S}}$ in a Bayesian detection framework. According to [31], the average KL distance obtained in (51) has the same expression as the Chernoff information, which determines the Bayesian detection error exponent. Therefore, the average KL distance measure in a sense maximizes the error exponent in the exponential decay on the Bayesian detection error performance (or miss detection performance under the Neyman-Pearson detection framework).

We note that it is possible that specific choices of the number of samplers P and the dictionary $\{\phi_{i,k,q}(t)\}_{i=1,\cdots,I}^{k\in\mathcal{K},q\in\mathcal{Q}}$ (i.e., the Gram matrix $\mathbf{M}$) lead to indistinguishable sparsity patterns [29] such that $\text{spark}\left(\mathcal{G}(\mathbf{B})\right) \leq |\mathcal{S}| + |\mathcal{S}'|$. In other words, the design of $\mathbf{B}$ cannot cure intrinsic problems caused by the choice of P or the Gram matrix $\mathbf{M}$, which are given parameters in the optimization. The intrinsic problems caused by the Gram matrix $\mathbf{M}$ in communications is typically handled by optimizing the transmit sequences $\phi_i(t)$ irrespective of the receiver choice such that $\mathbf{M}$ becomes diagonally dominated. This corresponds to having a well localized ambiguity function for each of the $\phi_i(t)$ and low cross-correlation between $\phi_i(t)$'s with different delays and Doppler. Gold sequences used in GPS and M-sequences used in spread spectrum communications, for example, are known to have good properties in this regard. This is a well investigated problem [32] that we do not aim to cover in this paper.

Given P and $\mathbf{M}$, we propose an optimal $\mathbf{B}$ that maximizes the average KL distance $\overline{\mathbb{D}}$ if there is a unique solution to the optimization; when there are multiple solutions that yield identical average KL distance $\overline{\mathbb{D}}$, we further choose in the feasible set the matrix $\mathbf{B}$ that gives the least occurrence of events $\mathbb{D}\left(\mathcal{H}_{\mathcal{S}} \| \mathcal{H}_{\mathcal{S}'}\right) = 0$. We use the results in the following lemma for our optimization.

Lemma 1. (Ratio Trace Maximization [33]) *Given two $L \times L$ positive semi-definite matrices $\mathbf{S}$ and $\mathbf{G}$, and an arbitrary $L \times P$ full column rank matrix $\mathbf{W}$, the ratio trace problem is formulated as*

$$\mathbf{W}^{\text{opt}} = \arg\max_{\mathbf{W}} \ \text{Tr}\left[\left(\mathbf{W}^H\mathbf{S}\mathbf{W}\right)^{-1}\mathbf{W}^H\mathbf{G}\mathbf{W}\right]. \tag{52}$$

The optimal $\mathbf{W}^{\text{opt}} = [\mathbf{w}_1^{\text{opt}}, \cdots, \mathbf{w}_P^{\text{opt}}]$ is given by the generalized eigenvectors $\mathbf{w}_p^{\text{opt}}$, $p = 1, \cdots, P$ corresponding to P largest generalized eigenvalues of the pair $(\mathbf{S}, \mathbf{G})$ with $P \leq \text{rank}[\mathbf{S}]$.

The optimal $\mathbf{B}$ is identified in the following theorem[3]:

[3]Note that the computation of the weights is done offline, and does not add complexity to the online processing.

Theorem 3. *Let* $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^H$*, where* $\mathbf{\Sigma} = \mathrm{diag}[\sigma_1, \cdots, \sigma_{I|\mathcal{K}||\mathcal{Q}|}]$ *is the eigenvalue matrix in descending order and* $\mathbf{U}$ *is the eigenvector matrix of* $\mathbf{M}$*. Denote the set of* $P \leq \mathrm{rank}[\mathbf{S}]$ *principal eigenvectors*

$$\mathcal{U} = \Big\{ \mathbf{U}_P = [\mathbf{u}_1, \cdots, \mathbf{u}_P] : \tag{53}$$

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^H, \mathbf{U} = \big[\mathbf{u}_1, \cdots, \mathbf{u}_{I|\mathcal{K}||\mathcal{Q}|}\big] \Big\}. \tag{54}$$

Let $\mathbf{\Xi}_P$ *be an arbitrary non-singular* $P \times P$ *matrix. When the eigenvectors are unique, the matrix* $\mathbf{B}_\star = \mathbf{\Xi}_P \mathbf{U}_P^H$ *is chosen uniquely to maximize the average KL distance* $\overline{\mathbb{D}}$ *in* (50)*. When the eigenvectors are not unique, we choose* $\widehat{\mathbf{U}}_P = \max_{\mathbf{U}_P \in \mathcal{U}} \mathrm{spark}\left(\mathbf{U}_P^H\right)$ *to maximize the average KL distance and minimize the occurrence of events* $\mathbb{D}\left(\mathcal{H}_\mathcal{S} \| \mathcal{H}_{\mathcal{S}'}\right) = 0$ *for* $\mathcal{S} \neq \mathcal{S}'$*.*

Proof: See Appendix D. ∎

The theorem above suggests that given $\mathbf{B}_\star$, the sampling kernels can be designed as

$$\psi_p(t) = \sum_{i=1}^{I} \sum_{k \in \mathcal{K}} \sum_{q \in \mathcal{Q}} b^\star_{p,(i,k,q)} \phi_{i,k,q}(t), \quad p = 1, \cdots, P. \tag{55}$$

Note that, as long as the preamble sequences do not change, the optimal matrix $\mathbf{B}_\star$ and the corresponding sampling kernels $\psi_p(t)$ are pre-computed only once and their design does not contribute to the running cost of the receiver operations. If the projections on the sampling kernels are implemented in the digital domain instead of being analog filters, then the samples of $\psi_p(t)$ are placed in the static memory that contains the receiver signal processing algorithms.

If the principal eigenvectors $\mathbf{U}_P$ are unique, then the choice of $\mathbf{B}_\star$ in general spreads out the pairwise KL distances, but $\mathbb{D}\left(\mathcal{H}_\mathcal{S} \| \mathcal{H}_{\mathcal{S}'}\right) = 0$ is possible for some choice of $\mathcal{S}$ and $\mathcal{S}'$. If $\mathrm{spark}(\mathbf{M}) \geq 2|\mathcal{I}|R$ and we choose $P = \mathrm{rank}[\mathbf{S}]$, then it can be ensured that $\overline{\mathbb{D}}$ is maximized and $\mathbb{D}\left(\mathcal{H}_\mathcal{S} \| \mathcal{H}_{\mathcal{S}'}\right) > 0$ is guaranteed for $\mathcal{S} \neq \mathcal{S}'$ with $|\mathcal{S}|, |\mathcal{S}'| \leq |\mathcal{I}|R$. On the other hand, an extreme example where the eigenvectors are not unique is when $\{\phi_{i,k,q}(t)\}_{i=1,\cdots,I}^{k \in \mathcal{K}, q \in \mathcal{Q}}$ form an orthogonal basis such that $\mathbf{M} = \mathbf{I}$. In this case, Theorem 3 is analogous to the fundamental criterion in compressed sensing that aims to find a matrix with $\mathrm{spark}(\mathbf{U}_P^H) \geq 2|\mathcal{I}|R$ that guarantees the recovery of any $|\mathcal{I}|R$-sparse vectors.

Remark: The number $|\mathcal{U}|$ of possible eigen-decompositions of the given matrix $\mathbf{M}$ may be quite large. For the extreme case when $\mathbf{M} = \mathbf{I}$, an arbitrary unitary matrix will be a possible choice. Fortunately, it is well known in compressive sensing that partial unitary matrices (such as the partial DFT matrix) have good compressive sensing properties (mutual coherence), thus this would not entail much loss if the matrix does not have exactly the maximum spark. On the other hand, as long as the number $|\mathcal{U}|$ is small, a finite search is also possible. More importantly, this task only needs to be done once and off-line.

VII. Numerical Results

In this section, we compare the C-SA acquisition scheme using the CMS architecture we propose against the alternative schemes. The first alternative is the D-SA scheme discussed in Section III-C, which processes the uncompressed Nyquist-rate samples $\mathbf{c}_{\text{DS}}[n]$ using sparsity recovery methods. Another alternative is the MF receiver, which also processes uncompressed samples. Rather than exploiting the underlying sparsity of the signal, it uses a filterbank matching the signal with all possible templates considered as hypotheses, trying to identify the link parameters through the best (highest) match.

To benchmark our C-SA against the D-SA and MF schemes we simulate the link acquisition of a single receiver plugged in a network populated by $I = 10$ users, out of which $|\mathcal{I}| = 4$ are randomly chosen to be actively transmitting. The user signature codes $\{a_i[m]\}$ in (1) belong to a set of M-sequences [34], which are quasi-orthogonal BPSK sequences of length $M = 255$ with unit power ($\|a_i[m]\|^2 = 1$). Due to the user dislocation, mobility and possible scattering, each path of this asynchronous multi-user channel is characterized by the triplet $\{h_{i,r}, t_{i,r}, \omega_{i,r}\}_{i \in \mathcal{I}, r \in \{1,\ldots,R\}}$ where $\{h_{i,r}\}$ are Rayleigh distributed, $h_{i,r} \sim \mathcal{CN}\big(0, 1/(R|\mathcal{I}|)\big)$, and uncorrelated, $\mathbb{E}\{h_{i,r} h^*_{i',r'}\} = 1/(|\mathcal{I}|R)\delta[i - i']\delta[r - r']$, normalized fading coefficients. The random delays $\{t_{i,r}\}$ are the sum of: (i) a time of arrival $t_0 = \min_{i,r} t_{i,r}$ that is uniformly distributed over an interval that spans the duration of the preamble $t_0 \in \mathcal{U}(0, MT)$, and of (ii) multipath delays that are uniformly distributed within an interval $(t_0, t_0 + t_{\max})$ where $t_{\max} \geq \max_{i,r,i',r'} |t_{i,r} - t_{i',r'}|$ is the maximum multipath delay spread of the channel. Consequently, all the arrival times are within a window of duration $2MT + t_{\max}$. The random frequency offsets $\{\omega_{i,r}\}$ are uniformly distributed, $\omega_{i,r} \in \mathcal{U}(-\omega_{\max}, \omega_{\max})$, over a range delimited at each side by the maximum Doppler spread $\omega_{\max}$. As we simulate underspread channel conditions, we choose $\omega_{\max}$ such that $\omega_{\max} t_{\max} \ll 2\pi$. Thus, for a multipath delay spread of $t_{\max} = 4T$ the choice of $\omega_{\max} = 2.5 \cdot 10^{-3} \times 2\pi/T$ is comparable to a 25 kHz offset for a 1 MHz signal.

More specifically, we compare the C-SA with the D-SA and the MF at the same resolutions for both Doppler and delays, which are, respectively, $\Delta\omega = \omega_{\max}/5 = 0.5 \times 2\pi/T$ and $\Delta\tau = T/2$. We test the C-SA using three different numbers of sampling channels $P = (60, 80, 100)$. On the other hand, the D-SA scheme uses Nyquist-rate samples per shift, which corresponds to using the whole spreading code duration of 255 samples per symbol, while the MF scheme uses a $I|\mathcal{K}| = 110$-channel filterbank and performs $I|\mathcal{K}||\mathcal{Q}| = 2640$ projections per shift. In fact, in the simulation, the sequential processing is performed by generating $N = D/T_s$ new Nyquist samples for each shift $t = nD$ with $D = 10T$, updating an internal buffer of size $W \leq \big((M + 2L_g)T + \tau_{\max}\big)/T_s$, where $L_g T = 3$ is the duration of the side lobe of $g(t)$ in (1) in samples. For a channel with a multipath delay spread of $t_{\max} = 4T$ and a shift of size $D = 10T$, the delay search space $\mathcal{Q}$ accounts for a potential displacement of $\tau_{\max} \geq t_{\max} + D = 14T$, therefore $\tau_{\max}/\Delta\tau \leq Q = 28$. For a frequency shift that spans the range $(-\omega_{\max}, \omega_{\max})$ the size of the discretized Doppler space is $K = 2(\omega_{\max}/\Delta\omega) + 1 = 11$. This value, together with the number of discretized delays, lead to a multi-user time-frequency grid of $I|\mathcal{K}||\mathcal{Q}| = 2640$ elements, common

for both C-SA and the alternative MF and D-SA schemes.

In the simulations, the C-SA compressive samples are generated from the same Nyquist-rate samples used for the other receivers, by projecting them onto the digitized version of the sampling kernels $\psi_p(nT_s)$, $p = 1, \ldots, P$, $n = 0, \ldots, M-1$. The C-SA simulation recovers the link parameters by solving (41) with the OMP algorithm [28], which is a popular choice to approximate the solution of a sparse problem [35]. To motivate our selection of OMP, we refer to Section VIII for an empirical evaluation of the OMP against two well-known ℓ_1 minimizers, SpaRSA [36] and ℓ_1-Homotopy [37], [38].

A. Signal Detection Performance

The first test on the detection performance of the C-SA receiver against the MF and the D-SA receiver we show, is for the cases of completely unknown active user sets, as discussed in Section V-B1. In Fig. 2, all receivers are unaware of the random set $\mathcal{I}$ of active users. Specifically, receivers consider as active components those that are found to have a signal strength that is at least 30% of the strongest components they estimate, i.e. in (44) $\rho_i = \max_{k,q} |\widehat{\beta}_{i,k,q}|^2 / 3$. If no possible component meets this requirement, the channel is declared idle. Events for the winning hypothesis $\mathcal{H}_{\widehat{\mathcal{A}}_{\ell^\star}}$ are generated according to (48). To first compare the sensitivity of the different receivers to active components, we define a signal hypothesis $\mathcal{H}_1$ corresponding to the all the non-idle channel hypotheses, i.e. $\widehat{\mathcal{A}}_{\ell^\star} \neq \varnothing$. Then, the detection sensitivity is measured in terms of the receiver operating characteristic (ROC) curve, tracing the probability of detection $P_d(\eta_0) = P\big(\eta_{\text{C-SA}}(\ell_\star) \geq \eta_0 | \mathcal{H}_1\big)$, against the probability of false alarm $P_f(\eta_0) = P\big(\eta_{\text{C-SA}}(\ell_\star) \geq \eta_0 | \mathcal{H}_\varnothing\big)$ when the channel is actually idle. Note that a positive detection may correspond to an incorrect identification of the specific users that are active. Thus, Section VII-B shows the rate of correct detection of active components for the same simulation scenario.

As it can be observed, although the C-SA receiver exploits less than $P/M = 80/255 \approx 1/3$ of the Nyquist-rate samples, the results from Fig. 2 show a modest degradation of the ROC compared to the MF receiver (less than 0.1 measured at $P_f(\eta_0) = 0.1$ and $\text{SNR} = -8$ dB). As expected, since the D-SA can leverage the additional observations to enhance its sensitivity, a growing gap $P_d(\eta_{\text{D-SA}}) - P_d(\eta_{\text{C-SA}})$ is observable as the SNR increases (measured at $P_f(\eta_0) = 0.1$) between the $\text{SNR} = -6$ dB and the $\text{SNR} = -12$ dB curves.

B. User Identification and Parameter Estimation Performance

In the simulations shown in Fig. 3(a) and 3(b), we measure the detection performance of $\widehat{\mathcal{I}}$ by the rate of successful identification $P(\widehat{\mathcal{I}} = \mathcal{I})$. In these simulations, the threshold η_0 is set to a level such that $P_f(\eta_0) \leq 0.1$, and thus we trace the curve at the point $P_f(\eta_0) = 0.1$ on the figures. The two sets of figures correspond to, respectively, the case where a receiver has partial knowledge of the active user components (as in for example the GPS receiver discussed in Section V-B2) and the exact same case examined previously in Fig. 2, where the receivers are unaware of the user $\mathcal{I}$. In the second case, for successful detection, not only the elements of the sets have

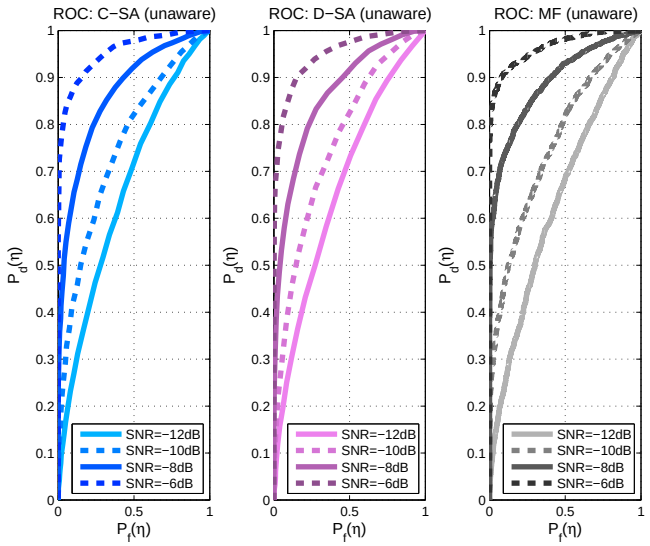
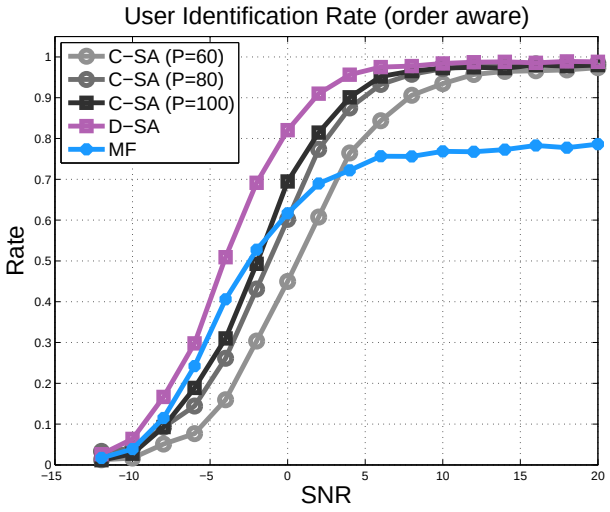


Fig. 2. Comparison of ROC curves for the order unaware receiver using the C-SA scheme with $P = 80$ channels (left), the D-SA scheme processing M uncompressed Nyquist observations (middle), and the MF scheme (right); tested at SNR=$\{-12, -10, -8, -6\}$ dB

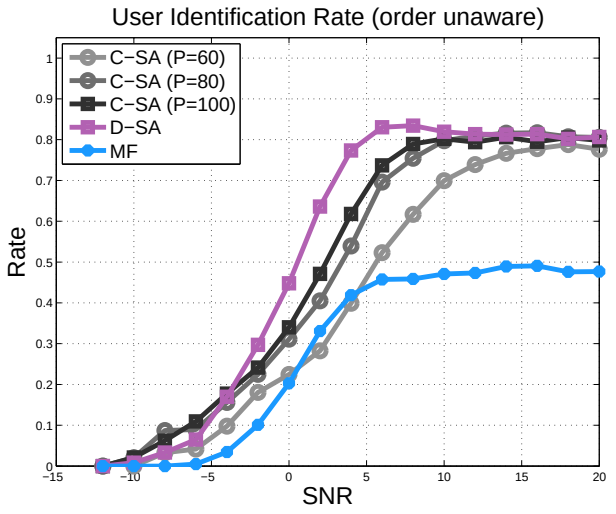
to be consistent $\widehat{\mathcal{I}} \subseteq \mathcal{I}$ but also their cardinality needs to be identical $|\widehat{\mathcal{I}}| = |\mathcal{I}|$. Instead for the first case, if the receiver has partial knowledge of the active components, then what matters is that the components are correctly identified, but their number is known ahead of time.

With a relatively short training sequence, we can see in Fig. 3(a) that the C-SA in the first case identifies the active user set with large probability (0.96 at $\text{SNR} = 20$ dB). The MF has worse performance due to the multi-user interference and to the presence of unresolvable paths. In fact, the MF receiver is unable to isolate the multi-path arrivals that fall within the same symbol period and, due to the presence of different Dopplers, its side-lobes may contribute negatively to the correlation, masking other active components. In contrast, the OMP algorithm in the C-SA scheme cancels the contributions from paths detected in previous iterations, before updating the projections to search for other components (the OMP processing steps are summarized in Section VIII). It is evident, however, that a low SNR, the C-SA scheme suffers from a loss due to the compression (-1 dB at the rate 0.6 with $P = 100$). This is clearly understood by observing the performance of the D-SA receiver as well. By processing uncompressed samples with the sparse reconstruction method, the D-SA curve combines the best of both worlds and, thus, its performance bounds the user identification rate for both the MF and the C-SA receivers in both examples. As shown in Fig. 3(b), the performance degrades when the receivers do not have side information on $|\mathcal{I}|$ (a difference of -0.13 for the CMS receiver against -0.3 of the MF at $\text{SNR} = 20$ dB). This is due to the cardinality mismatch, $\{|\widehat{\mathcal{I}}| \neq |\mathcal{I}|\}$, that occurs while estimating the order.

The accuracy of the recovered set $\widehat{\mathcal{A}}_{\ell^\star}$ is evaluated by the root mean square error (RMSE) of the parameters $\tau_{i,r}$ and $\omega_{i,r}$ that are associated to the correctly identified users $\mathcal{I} = \widehat{\mathcal{I}}$.

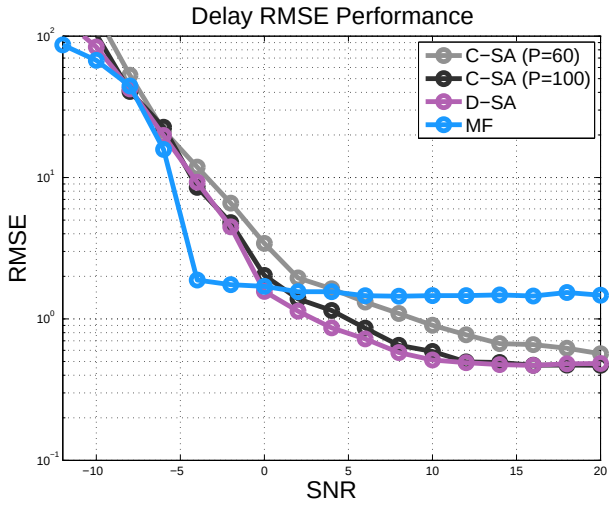


(a) Partial User Knowledge

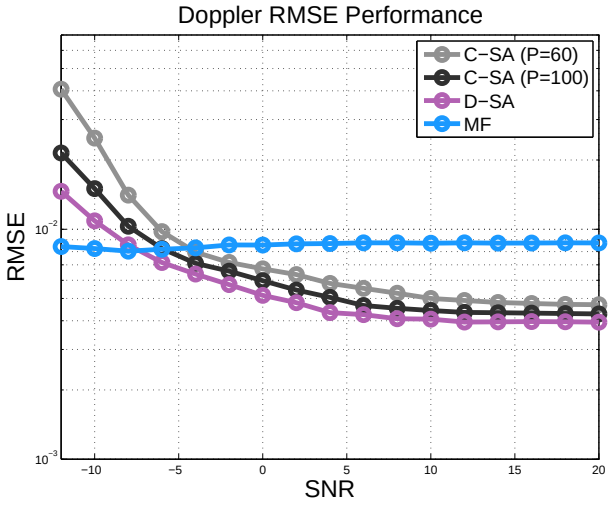


(b) No User Knowledge

Fig. 3. (a) Successful user identification rate $P(\widehat{\mathcal{I}} = \mathcal{I})$ for the order aware receiver implementing the MF (blue) receiver, the D-SA receiver (red), the CMS with C-SA (grey shades) receiver with $P = \{60, 80, 100\}$. (b) Successful user identification rate $P(\widehat{\mathcal{I}} = \mathcal{I})$ for the order unaware receiver implementing the MF (blue) receiver, the D-SA receiver (red), or the CMS (grey shades) receiver with $P = \{60, 80, 100\}$



(a) RMSE($\boldsymbol{\tau}$)



(b) RMSE($\boldsymbol{\omega}$)

Fig. 4. (a) RMSE($\boldsymbol{\tau}$) as a function of SNR $= \{-12, -8, \ldots, 20\}$ implemented by the C-SA scheme (grey shades) with $P = \{60, 100\}$, by the D-SA (red) or by the MF (blue). (b) RMSE($\boldsymbol{\omega}$) as a function of SNR $= \{-12, -8, \ldots, 20\}$ implemented by the C-SA scheme (grey shades) with $P = \{60, 100\}$, by the D-SA (red) or by the MF (blue).

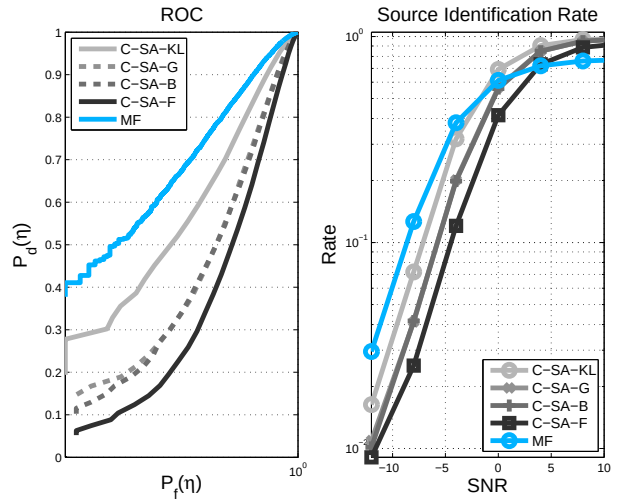
Thus,

$$\mathrm{RMSE}(\boldsymbol{\tau}) \triangleq \sqrt{\frac{1}{|\widehat{\mathcal{I}} \cap \mathcal{I}|} \frac{1}{R} \sum_{r=1}^{R} \sum_{i \in \{\widehat{\mathcal{I}} \cap \mathcal{I}\}} \left(\tau_{i,r} - \widehat{q}_{i,r} \Delta\tau\right)^2}$$

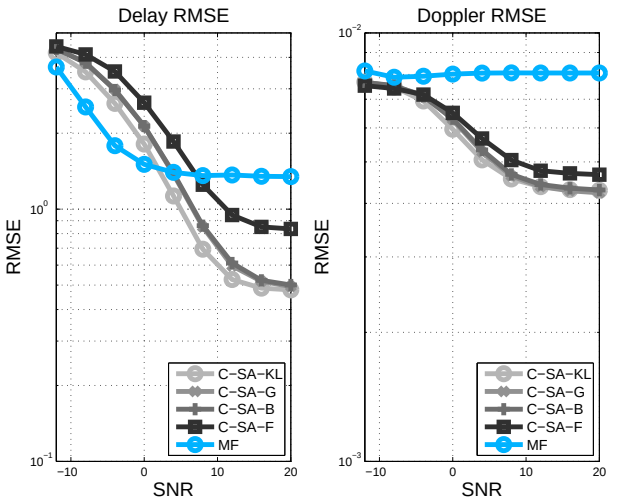
$$\mathrm{RMSE}(\boldsymbol{\omega}) \triangleq \sqrt{\frac{1}{|\widehat{\mathcal{I}} \cap \mathcal{I}|} \frac{1}{R} \sum_{r=1}^{R} \sum_{i \in \{\widehat{\mathcal{I}} \cap \mathcal{I}\}} \left(\omega_{i,r} - \widehat{k}_{i,r} \Delta\omega\right)^2}$$

are the average RMSE of the delay parameters the Doppler frequencies respectively.

To verify the accuracy of the parameter estimates of $\tau_{i,r}$ and $\omega_{i,r}$, we trace the RMSE's of the order-aware case. Once again we observe, from Fig. 4(a) and Fig. 4(b), that the performance of RMSE($\boldsymbol{\tau}$) and RMSE($\boldsymbol{\omega}$) is enhanced by the detector that better leverages the presence of the multi-path. In fact, at SNR $= 20$ dB, the accuracy of the CMS, with $P = 100$, and the D-SA approach the grid resolution, i.e. RMSE($\boldsymbol{\tau}$) $\approx \Delta\tau$ and RMSE($\boldsymbol{\omega}$) $\approx \Delta\omega$. Oppositely, the contribution of the unresolvable paths, in either frequency or time, to the correlation adversely affect the parameter selection. Not canceling the previously selected components contribute to a large error after the selection of the dominant paths as the same arrivals are likely to be selected more than once by the presence of correlated components. These errors impact the highest resolution since at SNR $= 20$ dB: RMSE($\boldsymbol{\tau}$) $> 2\Delta\tau$ and RMSE($\boldsymbol{\omega}$) $> 2\Delta\omega$. Instead, at low SNR, the performance is bounded by the maximum error given by the search space which is a function of $\omega_{\max}$ and $\tau_{\max}$, respectively, due to the early detection resulting from heavy noise.



(a) ROC and Identification Rate



(b) Parameter RMSE

Fig. 5. (a) ROC curve at SNR=−8 dB and user identification rate $P(\widehat{\mathcal{I}} = \mathcal{I})$ of the C-SA scheme, against the MF receiver and different choices of $\mathbf{B}$'s. (b) Delay and Doppler estimation RMSE of the C-SA scheme, against different random designs of $\mathbf{B}$ and the MF.

C. Optimality of Compressive Samplers

In this subsection, we briefly compare the performances of the C-SA scheme using a $P = 100$-channel CMS architecture with the optimal samplers versus other random projection schemes in compressed sensing. The ROC curve and the user identification rate $P(\widehat{\mathcal{I}} = \mathcal{I})$ in Fig. 5(a) show that the optimal sampling kernel, denoted by C-SA-KL, exhibits better performance than random designs of $\mathbf{B}$ using matrices whose entries are Gaussian (C-SA-G), Bernoulli (C-SA-B), or randomly selected rows of a DFT matrix (C-SA-F). It can also be observed from Fig. 5(b) that the RMSE of the delay and Doppler estimates are also improved.

D. Data Detection Performance after Link Acquisition

As a complementary evaluation of the proposed C-SA acquisition scheme, we provide here also the data detection performance in terms of Symbol Error Rate (SER), using

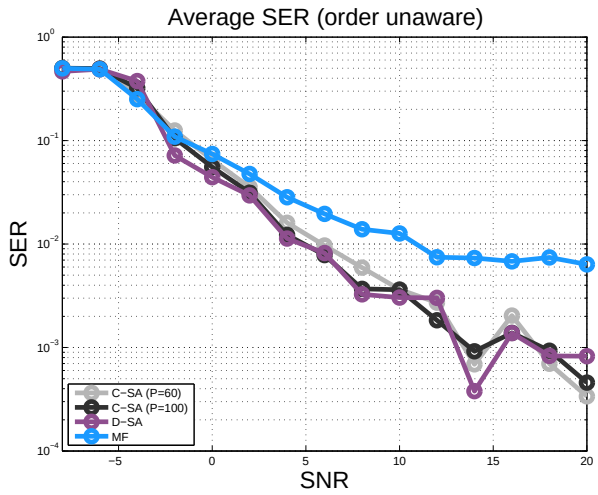


Fig. 6. SER performance of the CMS receiver with C-SA acquisition scheme against the D-SA and MF acquisition schemes.

linear minimum mean-square-error (LMMSE) multi-user receivers constructed by using the estimated signal spaces, and using the link parameters acquired by the C-SA, D-SA and MF link acquisition schemes respectively. For simplicity, we consider the scenario where the receivers have no knowledge of the active user components initially, as discussed in Section V-B1. We remark here that the SER performances of all the receivers are influenced consistently by the pre-defined grid resolution $\Delta\tau$ and $\Delta\omega$, therefore the error floors of parameter estimation observed in Fig. 4(a) and 4(b) ultimately result in an irreducible error floor on the SER curves for all the receivers. Thus, given a specific resolution, it can be seen in Fig. 6 that the parameters obtained by the C-SA scheme lead to a much better detection performance than that of the MF scheme, while closely approaching the performance of the D-SA scheme. If necessary, the resolution can be made finer such that the error floor decays to a satisfactory level according to the receiver requirement for each case, and other decoding strategies as well as decision directed and blind approaches can be combined to further improve performance. Once the signal subspace is coarsely identified, fine tuning and tracking the signal space and decoding symbols optimally becomes a rather standard problem. Hence this example is just aimed at illustrating the quality of the initialization to be expected from the architectures examined.

VIII. Cost Analysis

In this section, we explicitly analyze the implementation costs of the MF and the proposed C-SA architectures in terms of storage requirement and computational complexity. The cost benefits are illustrated in two regimes respectively: the *analog implementation*, that corresponds to what the paper describes mathematically in detail, and a *digital implementation*, which would be necessary if the projections on the compressive samplers $\psi_p(t)$ cannot be implemented in the form of analog filters. The compressive procedure in that case would emulate our implementation in simulations, where Nyquist samples of

the signal $x(t)$ are projected onto digital filters matched to the samples of $\psi_p(t)$.

The metric to evaluate *storage requirements* is given by the A/D hardware cost, measured as the number of sampling channels P which is the buffer size of the A/D samples, while the *computational complexity* is evaluated counting the *total amount of complex additions and multiplications*, and by the average CPU run time spent on executing tasks (a 64-bit i7 920 CPU running at 2.67 GHz). In the following, we first settle on the sparse recovery solver for the SR-LRT in the C-SA scheme, and then proceed with our comparison using the chosen solver.

A. Sparse Recovery Solver for the CMS Receiver: The OMP Algorithm

The C-SA receiver spends its greatest effort in solving the optimization (41). Fast ℓ_1 minimizers like SpaRSA [36] or ℓ_1-Homotopy [37], [38] are often the methods of choice. The former greatly reduces the complexity by approximating the Hessian of the gradient descent by a diagonal matrix, whereas the latter inverts a system of equations whose number of unknowns, at each iteration, remains restricted to the non-zero elements of the sparse vector estimate. Greedy algorithms like the OMP [28], are efficient approximations to the solution of sparse problems as well [35]. The OMP algorithm iteratively detects the strongest element in the sparse vector and removes its contribution in the next iteration; thus, the number of iterations required by OMP is bounded by the maximum possible components that exist in the signal, which in our case is $|\mathcal{I}|R$. The average CPU run time spent in solving (41) with different solvers is illustrated in Fig. 7(a) traced against P, where the OMP algorithm shows significantly less computation time. Thus, in Fig. 7(b) we further compare the average CPU run time of the CMS receiver using OMP against the MF receiver, the implementation details of which will be discussed in the following subsection. OMP has smaller complexity primarily because it stops as soon as all the strong entries have been detected. In contrast, ℓ_1-Homotopy and SpaRSA do not limit the search to a single set, but rather explore the feasible set by selecting and de-selecting elements of the support vector (ℓ_1-Homotopy), or by shrinking it through a gradient descent (SpaRSA), until a desired convergence criterion has been met.

B. C-SA Scheme v.s. MF Scheme

Using the OMP algorithm for sparse recovery, we summarize the steps of the C-SA and MF schemes in Algorithms 1 and 2 respectively.

Based on the algorithm descriptions, we provided the order of storage cost and computational complexities in the tables below. Storage accounts for a *data path storage* component, dynamically updated with the streaming data that correspond to new observations to be processed, and for a *static component*, that stores filters or sampling kernels parameters needed to perform signal processing on the data.

It is seen in Fig. 8 and 9 that both our C-SA scheme with the CMS receiver and the MF receiver (using exhaustive matched filtering) have computational complexities that scale linearly with the dimension of the search space $I|\mathcal{K}||\mathcal{Q}|$. However, the *data path storage* of the C-SA receivers is greatly reduced. Another storage gain is found in the case of digital implementation, because there are fewer projections to be made and thus, in principle, unless the MF are synthesized on the fly, a smaller amount of static memory is required to store the samples of $\psi_p(t)$.

When the architectures are implemented in the digital domain, the C-SA receiver also leads to a great reduction in computational complexity, with an approximate ratio with the MF receiver complexity of

$$\frac{MP + I|\mathcal{K}||\mathcal{Q}| + P^3}{MI|\mathcal{K}||\mathcal{Q}|} \approx \frac{P}{I|\mathcal{K}||\mathcal{Q}|} + \frac{P^3}{MI|\mathcal{K}||\mathcal{Q}|}. \tag{58}$$

Clearly, when the preamble sequence is long and the search space is large $MI|\mathcal{K}||\mathcal{Q}| \gg P^3$, this ratio becomes less than 1 and the C-SA architecture leads to computational savings while, as seen in our simulations, maintaining comparable performance. This is also why the C-SA receiver implemented in the simulation (see Fig. 7(b)) considerably outperforms the MF receiver in terms of average CPU run time for large $M \geq 3000$ with $P = 60, 80, 100$. On the other hand, when M is small (e.g. $M = 255$ in Section VII), the MF receiver has less computation time for $M < 3000$ against $P = 60, 80, 100$ as in Fig. 7(b), but such a short preamble

Algorithm 1 C-SA Scheme

(**CMS.1**) obtain compressive samples $\mathbf{c}[n]$ at the nth shift;

(**CMS.2**) initialize $\boldsymbol{\beta}^0 = \mathbf{0}$, $\mathcal{A}_n^0 = \varnothing$, $\bar{\mathbf{S}}_0 = \mathbf{BM}$, $\mathbf{S}_0 = \varnothing$, $j = 1$ and run the OMP algorithm;

(OMP.1) remove interference $\boldsymbol{\delta}^j = \mathbf{c}[n] - \mathbf{S}_{j-1}\left[\boldsymbol{\beta}^{j-1}\right]_{\mathcal{A}_n^{j-1}}$;

(OMP.2) projection $\boldsymbol{\xi}^j = \bar{\mathbf{S}}_{j-1}^T \boldsymbol{\delta}^j$, $\boldsymbol{\xi}^j = [\cdots, \xi_{i,k,q}^j, \cdots]^T$;

(OMP.3) detection $\mathcal{A}_n^j = \mathcal{A}_n^{j-1} \bigcup \{(i,k,q)\}$ with $(i,k,q) = \arg\max_{i,k,q} |\xi_{i,k,q}^j|^2$;

(OMP.4) update $\mathbf{S}_j = \left[\mathbf{BM}\right]_{(:,\mathcal{A}_n^j)}$ and $\bar{\mathbf{S}}_j = \left[\mathbf{BM}\right]_{(:,\overline{\mathcal{A}_n^j})}$;

(OMP.5) update the link vector

$$\left[\boldsymbol{\beta}^j\right]_{\mathcal{A}_n^j} = \left(\mathbf{S}_j^T \mathbf{S}_j\right)^{-1} \mathbf{S}_j^T \mathbf{c}[n] \tag{56}$$

$$\left[\boldsymbol{\beta}^j\right]_{\overline{\mathcal{A}_n^j}} = \mathbf{0}; \tag{57}$$

(OMP.6) stop if either $j = |\mathcal{I}|R$ or $\left\|\mathbf{c}[n] - \mathbf{BM}\boldsymbol{\beta}^j\right\| < \epsilon$, and set $j = j + 1$.

(**CMS.3**) Evaluate the likelihood ratio $\eta_{\text{C-SA}}(n)$ in (45) and check if it exceeds η_0.

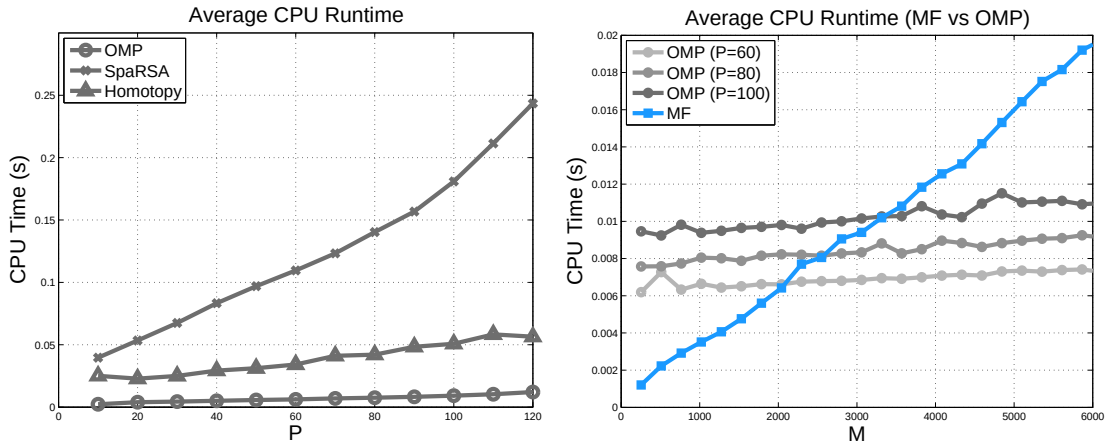
(**CMS.4**) If yes, then extract components accordingly (order-aware, order-unaware).

Algorithm 2 MF Scheme

(**MF.1**) obtain the sample array $\mathbf{C}_{\text{MF}}[n]$ in (17) from the MF filterbank;

(**MF.2**) identify the maximum output and check if it exceeds ρ_i for all $i = 1, \cdots, I$;

(**MF.3**) If yes, then extract the delay-Doppler set for each active user as in Section III.



(a) run time v.s. the number of sampling channels P

(b) run time v.s. the length of the preamble sequence M

Fig. 7. (a) Average CPU run time for the CMS receiver, implemented with different solvers (OMP, SpaRSA, ℓ_1-Homotopy), as a function of $P = \{10, 20, \cdots, 120\}$. (b) Average CPU runtime of a CMS receiver using $P = \{60, 80, 100\}$ compressed observations against the MF receiver. The average runtime is measured against the preamble length M.

CMS Receiver	**Storage** (analog)	**Complexity** (analog)	**Storage** (digital)	**Complexity** (digital)
(CMS.1)	$\mathcal{O}(P)$	0	$\mathcal{O}(MP)$	$\mathcal{O}(MP)$
(CMS.2)	0	$\mathcal{O}(I\|\mathcal{K}\|\|\mathcal{Q}\|) + \mathcal{O}(P^3)$	0	$\mathcal{O}(I\|\mathcal{K}\|\|\mathcal{Q}\|) + \mathcal{O}(P^3)$
(CMS.3)	$\mathcal{O}(P)$	$\mathcal{O}(P)$	$\mathcal{O}(P)$	$\mathcal{O}(P)$
Total	$\mathcal{O}(P)$	$\mathcal{O}(I\|\mathcal{K}\|\|\mathcal{Q}\|) + \mathcal{O}(P^3)$	$\mathcal{O}(MP)$	$\mathcal{O}(I\|\mathcal{K}\|\|\mathcal{Q}\|) + \mathcal{O}(P^3)$

Fig. 8. Complexity breakdown for the CMS receivers at every shift $t = nD$.

MF Receiver	**Storage** (analog)	**Complexity** (analog)	**Storage** (digital)	**Complexity** (digital)
(MF.1)	$\mathcal{O}(I\|\mathcal{K}\|\|\mathcal{Q}\|)$	0	$\mathcal{O}(MI\|\mathcal{K}\|\|\mathcal{Q}\|)$	$\mathcal{O}(MI\|\mathcal{K}\|\|\mathcal{Q}\|)$
(MF.2)	0	$\mathcal{O}(I\|\mathcal{K}\|\|\mathcal{Q}\|)$	0	$\mathcal{O}(I\|\mathcal{K}\|\|\mathcal{Q}\|)$
Total	$\mathcal{O}(I\|\mathcal{K}\|\|\mathcal{Q}\|)$	$\mathcal{O}(I\|\mathcal{K}\|\|\mathcal{Q}\|)$	$\mathcal{O}(MI\|\mathcal{K}\|\|\mathcal{Q}\|)$	$\mathcal{O}(MI\|\mathcal{K}\|\|\mathcal{Q}\|)$

Fig. 9. Complexity breakdown for the MF receivers at every shift $t = nD$.

does not provide sufficient processing gain for reliable link acquisition in the presence of multipath, as can be clearly seen from the numerical results (e.g., see Fig. 3(a)). Thus, when M is small, the gain of the C-SA receiver also lies in the superior acquisition performance demonstrated by the numerical results, except for the low SNR region where the C-SA is not sufficiently sensitive.

IX. Conclusions

In this paper, we proposed the SR-LRT receiver using a unified CMS architecture for link acquisition, which we refer to as the C-SA scheme. This scheme uses a sequential SR-LRT that jointly detects signal presence and recovers the active users with their link parameters. We optimized the CMS architecture to maximize the average Kullback-Leibler distance among the hypotheses tested in the SR-LRT and show that, with the optimal compressive samplers we propose, the receiver detection outperforms those with conventional compressed sensing alternatives. Furthermore, through the numerical comparison of the proposed architecture with the D-SA scheme and the MF approach, we have shown that the C-SA receiver can scale down its processing storage and complexity with greater flexibility, while maintaining satisfactory performance.

Appendix A
Proof of Theorem 1

Substituting (23) into (20), we have

$$c_p[n] = \sum_{i=1}^{I} \sum_{k\in\mathcal{K}} \sum_{q\in\mathcal{Q}} \alpha_{i,k,q} e^{\mathrm{i}k\Delta\omega\ell D} \left\langle \phi_{i,k,q}(t-\ell D), \psi_p(t-nD) \right\rangle + \left\langle v(t), \psi_p(t-nD) \right\rangle . \quad (59)$$

Define the $P \times I|\mathcal{K}||\mathcal{Q}|$ matrix

$$\left[\mathbf{M}_{\psi\phi}[n-\ell]\right]_{p,(i,k,q)} = R_{\psi_p\phi_{i,k,q}}\left[(n-\ell)D\right] \quad (60)$$

$$\triangleq \left\langle \phi_{i,k,q}(t-\ell D), \psi_p(t-nD) \right\rangle \quad (61)$$

and denote $v_p[n] \triangleq \langle v(t), \psi_p(t-nD)\rangle$ as the sample of filtered noise, whose covariance can be obtained as $\mathbb{E}\{v_p[n]v^*_{p'}[n]\} = \sigma^2 \langle \psi_p(t), \psi_{p'}(t)\rangle$ using $\mathbb{E}\{v(t)v^*(s)\} =$

$\sigma^2\delta(t-s)$. Therefore, the covariance matrix of the noise samples vector $\boldsymbol{\nu}[n] \triangleq [\nu_1[n], \cdots, \nu_P[n]]^T$ has a covariance matrix $\mathbf{R}_{vv} = \sigma^2\mathbf{R}_{\psi\psi}$ where $[\mathbf{R}_{\psi\psi}]_{p,p'} \triangleq \langle \psi_p(t), \psi_{p'}(t)\rangle$ is the Gram matrix of the kernels $\psi_p(t)$'s.

Denote $\mathbf{c}[n] \triangleq [c_1[n], \cdots, c_P[n]]^T$ as the length-P vector containing the samples acquired from the CMS filterbank at time $t = nD$. Given the link vector $\boldsymbol{\alpha}[\ell] = [\cdots, \alpha_{i,k,q}, \cdots]^T$ at the ℓth shift as (24), we then have the observation model in matrix form

$$\mathbf{c}[n] = \mathbf{M}_{\psi\phi}[n-\ell]\boldsymbol{\Gamma}[\ell]\boldsymbol{\alpha}[\ell] + \boldsymbol{\nu}[n]. \tag{62}$$

Using a sampling kernel constructed as

$$\psi_p(t) = \sum_{i=1}^{I}\sum_{k\in\mathcal{K}}\sum_{q\in\mathcal{Q}} b_{p,(i,k,q)}\phi_{i,k,q}(t), \tag{63}$$

the cross-correlation $R_{\psi_p\phi_{i,k,q}}[(n-\ell)D]$ in (60) can be evaluated as

$$R_{\psi_p\phi_{i,k,q}}[(n-\ell)D] \tag{64}$$

$$= \sum_{i'=1}^{I}\sum_{q'\in\mathcal{Q}}\sum_{k'\in\mathcal{K}} b_{p,(i',k',q')}R_{\phi_{i',k',q'}\phi_{i,k,q}}[(n-\ell)D], \tag{65}$$

where $R_{\phi_{i',k',q'}\phi_{i,k,q}}[(n-\ell)D] = \langle \phi_{i,k,q}(t-\ell D), \phi_{i',k',q'}(t-nD)\rangle$. With a change of variable $t' = t - nD - q'\Delta\tau$, the correlation can be written as

$$R_{\phi_{i',k',q'}\phi_{i,k,q}}[(n-\ell)D] = e^{\mathrm{i}k\Delta\omega(n-\ell)D}e^{-jk\Delta\omega q'\Delta\tau} \times R^{(k-k')}_{\phi_{i'},\phi_i}[(q'-q)\Delta\tau + (n-\ell)D], \tag{66}$$

where $R^{(k-k')}_{\phi_{i'},\phi_i}(\Delta t)$ is the ambiguity function

$$R^{(k-k')}_{\phi_{i'},\phi_i}(\Delta t) = \int \phi^*_{i'}(t)\phi_i(t-\Delta t)e^{-\mathrm{i}(k-k')\Delta\omega t}\mathrm{d}t. \tag{67}$$

From (64), $\mathbf{M}_{\psi\phi}[n-\ell]$ in (60) can be re-written as $\mathbf{M}_{\psi\phi}[n-\ell] = \mathbf{B}\mathbf{M}_{\phi\phi}[n-\ell]$, where

$$\Big[\mathbf{M}_{\phi\phi}[n-\ell]\Big]_{(i',k',q'),(i,k,q)} = R_{\phi_{i',k',q'}\phi_{i,k,q}}[(n-\ell)D]. \tag{68}$$

Then the observation model can be re-written as

$$\mathbf{c}[n] = \mathbf{B}\mathbf{M}_{\phi\phi}[n-\ell]\boldsymbol{\Gamma}[\ell]\boldsymbol{\alpha}[\ell] + \boldsymbol{\nu}[n]. \tag{69}$$

Finally, the Gram matrix of $\psi_p(t)$'s can be obtained accordingly as

$$\mathbf{R}_{\psi\psi} = \mathbf{B}\mathbf{M}_{\phi\phi}[0]\mathbf{B}^H, \tag{70}$$

which gives the noise covariance as $\mathbf{R}_{vv} = \sigma^2\mathbf{B}\mathbf{M}_{\phi\phi}[0]\mathbf{B}^H$.

Appendix B
Proof of Theorem 2

From (59), each sample $c_p[n]$ from the CMS architecture $p = 1, \cdots, P$ can be expressed as

$$c_p[n] = \sum_{i'=1}^{I}\sum_{k'\in\mathcal{K}}\sum_{q'\in\mathcal{Q}} b_{p,(i',k',q')}\sum_{i=1}^{I}\sum_{k\in\mathcal{K}}\sum_{q\in\mathcal{Q}}\alpha_{i,k,q}e^{\mathrm{i}k\Delta\omega\ell D} \times R_{\phi_{i',k',q'}\phi_{i,k,q}}[(n-\ell)D] + v_p[n]. \tag{71}$$

The summation $\sum_{q\in\mathcal{Q}}\alpha_{i,k,q}e^{\mathrm{i}k\Delta\omega\ell D}R_{\phi_{i',k',q'}\phi_{i,k,q}}[(n-\ell)D]$ can be adjusted with respect to the relative time index $[n-\ell]$ by re-writing the correlation in (66) as

$$R_{\phi_{i',k',q'}\phi_{i,k,q}}[(n-\ell)D] \tag{72}$$

$$= e^{\mathrm{i}k\Delta\omega(n-\ell)D}e^{-jk\Delta\omega q'\Delta\tau}R^{(k-k')}_{\phi_{i'},\phi_i}[(q'-q)\Delta\tau + (n-\ell)D]$$

$$= e^{\mathrm{i}k\Delta\omega(n-\ell)D}e^{-jk\Delta\omega q'\Delta\tau}R^{(k-k')}_{\phi_{i'},\phi_i}[(q'-[q-(n-\ell)N])\Delta\tau]$$

$$= e^{\mathrm{i}k\Delta\omega(n-\ell)D}R_{\phi_{i',k',q'}\phi_{i,k,q-(n-\ell)N}}[0]. \tag{73}$$

Without loss of generality, let $D/\Delta\tau = N \in \mathbb{Z}$. With a change of variable $q'' = q-(n-\ell)N$ and substituting the equivalent correlation in (72), we have

$$\sum_{q\in\mathcal{Q}}\alpha_{i,k,q}e^{\mathrm{i}k\Delta\omega\ell D}R_{\phi_{i',k',q'}\phi_{i,k,q}}[(n-\ell)D]$$

$$\overset{(72)}{=} \sum_{q\in\mathcal{Q}} e^{\mathrm{i}k\Delta\omega nD}\alpha_{i,k,q}R_{\phi_{i',k',q'}\phi_{i,k,q-(n-\ell)N}}[0]$$

$$= \sum_{q''\in\mathcal{Q}} e^{\mathrm{i}k\Delta\omega nD}\alpha_{i,k,q''+(n-\ell)N}R_{\phi_{i',k',q'}\phi_{i,k,q''}}[0],\ q'' = q-(n-\ell)N.$$

With the re-formulation, (71) is re-written as below

$$c_p[n] = \sum_{i'=1}^{I}\sum_{k'\in\mathcal{K}}\sum_{q'\in\mathcal{Q}} b_{p,(i',k',q')}\sum_{i=1}^{I}\sum_{k\in\mathcal{K}}\sum_{q\in\mathcal{Q}} e^{\mathrm{i}k\Delta\omega nD}\alpha_{i,k,q+(n-\ell)N} \times R_{\phi_{i',k',q'}\phi_{i,k,q}}[0] + v_p[n].$$

By letting $\mathbf{M} \triangleq \mathbf{M}_{\phi\phi}[0]$ and defining the shifted link vector $\boldsymbol{\alpha}[n]$ at the nth shift as

$$\Big[\boldsymbol{\alpha}[n]\Big]_{(i,k,q)} \triangleq \alpha_{i,k,q+(n-\ell)N}, \tag{74}$$

the observation model can be equivalently re-written as

$$\mathbf{c}[n] = \mathbf{B}\mathbf{M}\boldsymbol{\Gamma}[n]\boldsymbol{\alpha}[n] + \boldsymbol{\nu}[n]. \tag{75}$$

Appendix C
Proof of Proposition 1

The pair-wise KL distance in (49) can be re-written with the trace operator $\mathrm{Tr}(\cdot)$ below

$$\mathbb{D}\left(\mathcal{H}_{\mathcal{S}}\|\mathcal{H}_{\mathcal{S}'}\right) = \frac{1}{\sigma^2}\mathrm{Tr}\left[\mathbf{M}^H\mathbf{B}^H\left(\mathbf{B}\mathbf{M}\mathbf{B}^H\right)^{-1}\mathbf{B}\mathbf{M}\mathbf{R}_{\boldsymbol{\beta}_{\mathcal{S}},\boldsymbol{\beta}_{\mathcal{S}'}}\right],$$

where $\mathbf{R}_{\boldsymbol{\beta}_{\mathcal{S}},\boldsymbol{\beta}_{\mathcal{S}'}} \triangleq (\boldsymbol{\beta}_{\mathcal{S}} - \boldsymbol{\beta}_{\mathcal{S}'})(\boldsymbol{\beta}_{\mathcal{S}} - \boldsymbol{\beta}_{\mathcal{S}'})^H$. Then the average pair-wise KL distance $\overline{\mathbb{D}}$ in (50) becomes

$$\overline{\mathbb{D}} = \frac{1}{\sigma^2}\mathrm{Tr}\left[\mathbf{M}^H\mathbf{B}^H\left(\mathbf{B}\mathbf{M}\mathbf{B}^H\right)^{-1}\mathbf{B}\mathbf{M}\mathbf{R}\right],$$

where $\mathbf{R} \triangleq \sum_{\mathcal{S}}\sum_{\mathcal{S}'}\gamma_{\mathcal{S},\mathcal{S}'}\mathbf{R}_{\mathcal{S},\mathcal{S}'}$ and $\mathbf{R}_{\mathcal{S},\mathcal{S}'}$ is the averaged covariance matrix of $\boldsymbol{\beta}_{\mathcal{S}}$ over the amplitudes

$$\mathbf{R}_{\mathcal{S},\mathcal{S}'} = \iint P(\boldsymbol{\beta}_{\mathcal{S}})P(\boldsymbol{\beta}_{\mathcal{S}'})\mathbf{R}_{\boldsymbol{\beta}_{\mathcal{S}},\boldsymbol{\beta}_{\mathcal{S}'}}\mathrm{d}\boldsymbol{\beta}_{\mathcal{S}}\mathrm{d}\boldsymbol{\beta}_{\mathcal{S}'}.$$

Given $P(\boldsymbol{\beta}_{\mathcal{S}}) = \prod_{(i,k,q)\in\mathcal{S}}P(\beta_{i,k,q})$ with $\int\boldsymbol{\beta}_{\mathcal{S}}P(\boldsymbol{\beta}_{\mathcal{S}})\mathrm{d}\boldsymbol{\beta}_{\mathcal{S}} = \mathbf{0}$ and $\int|\beta_{i,k,q}|^2P(\beta_{i,k,q})\mathrm{d}\beta_{i,k,q} = $ constant, the averaged matrix $\mathbf{R}_{\mathcal{S},\mathcal{S}'}$ is diagonal. Furthermore, if the set of weights $\gamma_{\mathcal{S},\mathcal{S}'}$ are constant for all $\mathcal{S},\mathcal{S}'$ and the individual weighting function $P(\beta_{i,k,q})$ is identical for all i,k,q, it also satisfies $\mathbf{R} \propto \mathbf{I}$ because the summation over $\mathcal{S},\mathcal{S}'$ is symmetric, and hence produces equal sum. Thus the result follows.

APPENDIX D
PROOF OF THEOREM 3

By analogy with Lemma 1, we have $\mathbf{S} = \mathbf{M}$ and $\mathbf{G} = \mathbf{M}\mathbf{M}^H$ in (51). Let $\mathbf{B} \triangleq \begin{bmatrix}\mathbf{b}_1 & \cdots & \mathbf{b}_P\end{bmatrix}^H$, where $\mathbf{b}_p$ is a length-$I|\mathcal{K}||\mathcal{Q}|$ column vector such that $\mathbf{b}_p = \mathbf{w}_p$. In this setting, according to Lemma 1, the optimal $\mathbf{b}_p$ is chosen as the generalized eigenvector of the matrix pair $(\mathbf{S}, \mathbf{G})$ such that $\mathbf{M}\mathbf{b}_p = \lambda_p \mathbf{M}\mathbf{M}^H\mathbf{b}_p$. Using the eigen-decomposition of $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^H$ and the property $\mathbf{U}^H\mathbf{U} = \mathbf{I}$, we have

$$\mathbf{\Sigma}\mathbf{U}^H\mathbf{b}_p = \lambda_p \mathbf{\Sigma}\mathbf{\Sigma}^H\mathbf{U}^H\mathbf{b}_p, \quad p = 1, \cdots, P. \tag{76}$$

If we choose $\mathbf{b}_p = \mathbf{u}_p$, where $\mathbf{u}_p$ is the pth column in the matrix $\mathbf{U}$, then the above relationship holds for all $p = 1, \cdots, P$ as long as $P \leq \mathrm{rank}(\mathbf{\Sigma})$ because $\mathbf{u}_i^H\mathbf{u}_j = \delta[i-j]$. This gives

$$\begin{aligned}\text{L.H.S.} \ &: \ \sigma_p \mathbf{U}^H\mathbf{u}_p = \sigma_p \mathbf{e}_p, \\ \text{R.H.S.} \ &: \ \lambda_p \mathbf{\Sigma}\mathbf{\Sigma}^H\mathbf{U}^H\mathbf{u}_p = \lambda_p \sigma_p^2 \mathbf{e}_p,\end{aligned}$$

leading to a generalized eigenvalue of $\lambda_p = 1/\sigma_p$, where $\sigma_p > 0$ is the pth eigenvalue in $\mathbf{\Sigma}$ and $\mathbf{e}_p$ is the canonical basis with 1 in the pth entry and 0 otherwise. Denote by $\mathbf{\Sigma}_P$ and $\mathbf{U}_P$ the principal eigenvalue and eigenvector matrices. Then the optimal $\mathbf{B}$ is chosen as $\mathbf{B} = \mathbf{\Xi}_P \mathbf{U}_P^H$, where $\mathbf{\Xi}_P$ is an arbitrary non-singular $P \times P$ matrix. According to (51), this choice gives

$$\overline{\mathbb{D}} = \mathrm{Tr}\left(\mathbf{\Sigma}_P^H \underbrace{\mathbf{\Xi}_P^H \mathbf{\Xi}_P^{-H}}_{=\mathbf{I}} \mathbf{\Sigma}_P^{-1} \underbrace{\mathbf{\Xi}_P^{-1} \mathbf{\Xi}_P}_{=\mathbf{I}} \mathbf{\Sigma}_P \right) = \sum_{p=1}^{P} \sigma_p,$$

which is independent of $\mathbf{\Xi}_P$. If the principal eigenvectors $\mathbf{U}_P$ are unique, the above $\mathbf{B}$ uniquely maximizes the average KL distance $\overline{\mathbb{D}}$. This choice of $\mathbf{B}$ in general spreads out the individual KL distance, while the occurence of the events $\mathbb{D}(\mathcal{H}_\mathcal{S} \| \mathcal{H}_{\mathcal{S}'}) = 0$ is analyzed below. So is the case when $\mathbf{U}_P$ is not unique.

Now we examine the occurence of $\mathbb{D}(\mathcal{H}_\mathcal{S} \| \mathcal{H}_{\mathcal{S}'}) = 0$. Let $\boldsymbol{\beta}_{\mathcal{S}\cup\mathcal{S}'} = (\boldsymbol{\beta}_\mathcal{S} - \boldsymbol{\beta}_{\mathcal{S}'})$ be a sparse vector with $|\mathcal{S}|, |\mathcal{S}'| \leq s$, and $s \leq |\mathcal{I}|R$. Substituting the matrix $\mathbf{B} = \mathbf{\Xi}_P \mathbf{U}_P^H$ back to (49) and simplifying the expression, the individual KL distance is

$$\mathbb{D}(\mathcal{H}_\mathcal{S} \| \mathcal{H}_{\mathcal{S}'}) = \frac{1}{\sigma^2} \boldsymbol{\beta}_{\mathcal{S}\cup\mathcal{S}'}^H \mathbf{U}_P \mathbf{\Sigma}_P \mathbf{U}_P^H \boldsymbol{\beta}_{\mathcal{S}\cup\mathcal{S}'}, \tag{77}$$

$$\forall \mathcal{S} \neq \mathcal{S}', \ |\mathcal{S}|, |\mathcal{S}'| \leq s. \tag{78}$$

Note that $\boldsymbol{\beta}_{\mathcal{S}\cup\mathcal{S}'}$ is a $2s$-sparse vector and $\mathbb{D}(\mathcal{H}_\mathcal{S} \| \mathcal{H}_{\mathcal{S}'})$ is bounded away from zero as long as any $2s$-sparse vectors do not fall into the null space of the matrix $\mathbf{U}_P^H$. In order to minimize the occurrence of the event $\mathbb{D}(\mathcal{H}_\mathcal{S} \| \mathcal{H}_{\mathcal{S}'}) = 0$, it is equivalent to maximizing the kruskal rank of the matrix $\mathbf{U}_P^H$ such that the matrix $\mathbf{B}$ can recover any s-sparse vector $\boldsymbol{\beta}_\mathcal{S}$ with s being maximized in this process.

REFERENCES

[1] A. Fletcher, S. Rangan, and V. Goyal, "On-Off Random Access Channels: A Compressed Sensing Framework," *Arxiv preprint ArXiv:0903.1022*, 2009.

[2] L. Applebaum, W. Bajwa, M. Duarte, and R. Calderbank, "Asynchronous Code-Division Random Access using Convex Optimization," *Physical Communication*, 2011.

[3] S. Verdu, *Multiuser Detection*. Cambridge Univ Pr, 1998.

[4] Y. Xie, Y. C. Eldar, and A. Goldsmith, "Reduced-Dimension Multiuser Detection," in *submitted to Information Theory, IEEE Transactions on*. IEEE, 2011.

[5] B. Sirkeci-Mergen and A. Scaglione, "Signal Acquisition for Cooperative Transmissions in Multi-hop Ad-hoc Networks," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, vol. 2. IEEE, 2004, pp. ii–345.

[6] S. Buzzi, A. De Maio, and M. Lops, "Code-Aided Blind Adaptive New User Detection in DS/CDMA Systems with Fading Time-Dispersive Channels," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2637–2649, 2003.

[7] L. Qiu, Y. Huang, and J. Zhu, "Fast Acquisition Scheme and Implementation of PRACH in WCDMA System," in *Vehicular Technology Conference, 2001. VTC 2001 Fall. IEEE VTS 54th*, vol. 3. IEEE, 2001, pp. 1701–1705.

[8] R. Wang and H. Li, "Decorrelating Multiuser Code-Timing Estimation for Long-Code CDMA with Bandlimited Chip Waveforms," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2369–2381, 2005.

[9] E. Fishler, A. Haimovich, R. Blum, L. Cimini Jr, D. Chizhik, and R. Valenzuela, "Spatial Diversity in RADARS-Models and Detection Performance," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 823–838, 2006.

[10] Z. Tian and G. Giannakis, "A GLRT Approach to Data-Aided Timing Acquisition in UWB Radios-Part I: Algorithms," *Wireless Communications, IEEE Transactions on*, vol. 4, no. 6, pp. 2956–2967, 2005.

[11] H. Zhu and G. Giannakis, "Exploiting Sparse User Activity in Multiuser Detection," *Communications, IEEE Transactions on*, vol. 59, no. 2, pp. 454–465, 2011.

[12] M. Duarte, M. Davenport, M. Wakin, and R. Baraniuk, "Sparse Signal Detection from Incoherent Projections," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 3. IEEE, 2006, pp. III–III.

[13] M. Davenport, P. Boufounos, M. Wakin, and R. Baraniuk, "Signal Processing with Compressive Measurements," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 445–460, 2010.

[14] J. Haupt and R. Nowak, "Compressive Sampling for Signal Detection," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 3. IEEE, 2007, pp. III–1509.

[15] Z. Wang, G. Arce, and B. Sadler, "Subspace Compressive Detection for Sparse Signals," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 3873–3876.

[16] J. Paredes, Z. Wang, G. Arce, and B. Sadler, "Compressive Matched Subspace Detection," in *European Signal Processing Conf*, 2009.

[17] K. Gedalyahu and Y. C. Eldar, "Time-Delay Estimation from Low-Rate Samples: A Union of Subspaces Approach," *Signal Processing, IEEE Transactions on*, vol. 58, no. 6, pp. 3017–3031, 2010.

[18] W. Bajwa, K. Gedalyahu, and Y. Eldar, "Identification of parametric underspread linear systems and super-resolution radar," *Signal Processing, IEEE Transactions on*, no. 99, pp. 1–1, 2011.

[19] M. Vetterli, P. Marziliano, and T. Blu, "Sampling Signals with Finite Rate of Innovation," *Signal Processing, IEEE Transactions on*, vol. 50, no. 6, pp. 1417–1428, 2002.

[20] I. Maravic and M. Vetterli, "Sampling and Reconstruction of Signals with Finite Rate of Innovation in the Presence of Noise," *Signal Processing, IEEE Transactions on*, vol. 53, no. 8, pp. 2788–2805, 2005.

[21] K. Gedalyahu, R. Tur, and Y. C. Eldar, "Multichannel Sampling of Pulse Streams at the Rate of Innovation," *Signal Processing, IEEE Transactions on*, vol. 59, no. 4, pp. 1491–1504, 2011.

[22] J. Urigüen, Y. C. Eldar, P. Dragotti, and Z. Ben-Haim, "Sampling at the rate of innovation: Theory and applications," 2012.

[23] Y. C. Eldar, "Compressed Sensing of Analog Signals in Shift-Invariant Spaces," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 2986–2997, 2009.

[24] S. Qaisar and A. Dempster, "Cross-correlation Performance Assessment of Global Positioning System (gps) l_1 and l_2 Civil Codes for Signal Acquisition," *Radar, Sonar & Navigation, IET*, vol. 5, no. 3, pp. 195–203, 2011.

[25] J. Kim, S. Sarin, M. Yasunaga, and H. Oh, "Robust Noncoherent PN-code Acquisition for CDMA Communication Systems," *Vehicular Technology, IEEE Transactions on*, vol. 50, no. 1, pp. 278–286, 2001.

[26] S. Kay, "Fundamentals of Statistical Signal Processing : Volume I & II, 1993."

[27] X. Li, A. Rueetschi, Y. C. Eldar, and A. Scaglione, "GPS Signal Acquisition via Compressive Multichannel Sampling," *Physical Communication*, 2011.

[28] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. IEEE, 1993, pp. 40–44.

[29] E. Candes, J. Romberg, and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[30] T. Cover, J. Thomas, J. Wiley *et al.*, *Elements of Information Theory*. Wiley Online Library, 1991, vol. 6.

[31] D. Bajovic, B. Sinopoli, and J. Xavier, "Robust Linear Dimensionality Reduction for Hypothesis Testing with Application to Sensor Selection," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*. IEEE, 2009, pp. 363–370.

[32] J. Klauder, "The Design of Radar Signals Having both High Range Resolution and High Velocity Resolution," *Bell System Technical Journal*, vol. 39, pp. 809–820, 1960.

[33] R. Duda, P. Hart, and D. Stork, "Pattern Classification and Scene Analysis 2nd ed." 1995.

[34] G. Buracas and G. Boynton, "Efficient Design of Event-related fMRI Experiments using M-sequences," *Neuroimage*, vol. 16, no. 3, pp. 801–813, 2002.

[35] J. Tropp and A. Gilbert, "Signal Recovery from Random Measurements via Orthogonal Matching Pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, 2007.

[36] S. Wright, R. Nowak, and M. Figueiredo, "Sparse Reconstruction by Separable Approximation," *Signal Processing, IEEE Transactions on*, vol. 57, no. 7, pp. 2479–2493, 2009.

[37] M. Salman Asif and J. Romberg, "Dynamic Updating for Minimization," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 421–434, 2010.

[38] D. Malioutov, M. Cetin, and A. Willsky, "Homotopy Continuation for Sparse Signal Representation," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, vol. 5. IEEE, 2005, pp. v–733.